Dynamic Attention Behavior Under Return Predictability*

Daniel Andrei† Michael Hasler‡

May 3, 2017

Abstract

We investigate the dynamic problem of how much attention an investor should pay to news in order to learn about stock-return predictability and maximize expected lifetime utility. We show that the optimal amount of attention is U-shaped in the return predictor, increasing with both uncertainty and the magnitude of the predictive coefficient, and decreasing with stock-return volatility. The optimal risky asset position exhibits a negative hedging demand that is hump-shaped in the return predictor. Its magnitude is larger when uncertainty increases, but smaller when stock-return volatility increases. We test and find empirical support for these theoretical predictions.

*We would like to thank Patrick Augustin, Craig Boreth, Michael Brennan, Peter Christoffersen, Alexander David, Alexandre Jeanneret, Chayawat Ornthanalai, Alberto Rossi, and Alexey Rubtsov for insightful suggestions and comments. We are also grateful for comments received from conference/seminar participants at HEC-McGill Winter Finance Conference 2016, ISF 2015, QWAFAFEW 2016, Ryerson University, UCLA Anderson, and University of Toronto. Financial support from UCLA and the University of Toronto is gratefully acknowledged.

†UCLA, Anderson School of Management, 110 Westwood Plaza, Suite C420, Los Angeles, CA 90095, USA, daniel.andrei@anderson.ucla.edu; www.danielandrei.info

‡University of Toronto, Rotman School of Management, 105 St. George Street, Toronto, ON, M5S 3E6, Canada, Michael.Hasler@rotman.utoronto.ca; www.utoronto.hasler

1 Introduction

Growing empirical evidence suggests that investors' attention is time-varying.[1] This affects trading and asset prices, and therefore significantly impacts financial markets.[2] The aim of this paper is to show that these fluctuations result from a rational information gathering behavior.

We build a dynamic portfolio choice model to investigate the problem of how much attention an investor should pay to news. Our setup is based on the incomplete-information literature,[3] which implicitly assumes that the set of available information is exogenously given. The objective here is to relax this assumption and assume that information is optimally acquired at a cost. Similar to Huang and Liu (2007), we consider the joint portfolio choice and information acquisition problem faced by an investor who does not observe expected stock returns. While information acquisition is a one time choice in Huang and Liu (2007), it is dynamic in our framework. This allows us to provide empirically testable predictions on the dynamic relation between attention, risky investments, and the relevant state variables.[4]

In our theoretical model, an agent can invest in one risk-free asset and one risky stock with unobservable expected returns. At each point in time, the investor optimally chooses her consumption, portfolio, and quantity of information needed to estimate expected returns and maximize expected lifetime utility of consumption. Information acquisition regulates both the learning and the investment decisions of the investor. By acquiring more accurate information, i.e. by paying more *attention to news*,[5] the investor is able to better estimate expected returns and, therefore, to increase her expected utility, but at the expense of decreasing her current wealth. In other words, the investor faces a dynamic trade-off problem of asset and attention allocation.

The investor assumes that expected returns are a linear function of an observable predictive

[1] Barber and Odean (2008); Da, Engelberg, and Gao (2011); Sicherman, Loewenstein, Seppi, and Utkus (2015); Fisher, Martineau, and Sheng (2016).

[2] Chien, Cole, and Lustig (2012); Andrei and Hasler (2014); Fisher et al. (2016); Hasler and Ornthanalai (2017).

[3] Refer to the seminal papers by Detemple (1986); Gennotte (1986); Dothan and Feldman (1986).

[4] One related segment of the literature postulates that information has a hedonic impact on utility (Loewenstein, 1987; Caplin and Leahy, 2001, 2004; Brunnermeier and Parker, 2005; Pagel, 2013; Andries and Haddad, 2014). This can generate different levels of attention depending on the state of the world and can thus explain the fluctuations in attention we observe. Another part of the literature postulates that investors have limited ability to process information (Sims, 2003; Van Nieuwerburgh and Veldkamp, 2006, 2010; Kacperczyk, Van Nieuwerburgh, and Veldkamp, 2016). Therefore, information does not directly enter the utility function, but indirectly helps investors make better decisions under uncertainty.

[5] See Sims (2003), Peng and Xiong (2006), Kacperczyk et al. (2016), Mondria (2010), Mondria and Quintana-Domeque (2013) for a similar interpretation of attention.

variable (Xia, 2001).[6] The focus of the paper is on the predictive coefficient, which is stochastic and *unobservable*. The investor uses all of the available historical data to estimate the predictive coefficient and then construct a forecast of returns. In additon to historical data, the investor can also choose to improve her estimation by collecting news, but at a cost. Because perfect learning—observing expected returns—has an infinite cost, the investor must choose an optimal finite amount of attention. We characterize this optimal amount of attention and its responsiveness to changes in the state variables of the model.

The predictive variable determines, to a large extent, the optimal amount of attention to news. We show that the investor pays more attention to news the further away the predictive variable is from its long-term mean. In these states, the investor attempts to profit from the reversion to the mean of expected returns and thus acquires additional information to bet on the upcoming trend. Because this arises whenever the predictive variable is far away from its long-term mean, attention exhibits a U-shaped pattern.

Furthermore, we show that greater uncertainty unambiguously increases attention to news. The reason is that greater uncertainty increases the volatility of *expected returns* and therefore increases the likelihood of large future trends. Since the investor can efficiently exploit these trends only if her estimate of the predictive coefficient is accurate, the optimal decision is to pay more attention to news.

We also find that attention to news decreases with the volatility of *realized returns*, which we assume to be stochastic (Chacko and Viceira, 2005; Liu, 2007). When the volatility of realized returns increases, the quality of information provided by those returns deteriorates, which decreases the volatility of expected returns. This, in turn, decreases the likelihood of large future trends and prompts the investor to decrease her attention to news.[7]

Finally, attention to news increases with the magnitude of the estimated predictive coefficient. A predictive coefficient of large magnitude implies highly volatile expected returns, which prompts the investor to be particularly attentive to news.

[6] See also Kandel and Stambaugh (1996), Barberis (2000), and Brandt, Goyal, Santa-Clara, and Stroud (2005) for equivalent assumptions.

[7] Note that in our model there is a distinction between "expected return volatility" and "realized return volatility." The former depends on all state variables and can be thought of as a forward looking measure (which can be proxied by the VIX index), whereas the latter is a contemporaneous measure. As such, our model predicts that attention to news increases when the expected return volatility is high (see also Fisher et al., 2016) but decreases when the realized return volatility is high.

The optimal risky investment share increases with the Sharpe ratio of the stock, and features a negative hedging demand because expected returns are positively correlated to returns. The hedging demand is hump-shaped in the predictive variable, it is more negative when uncertainty increases, and it is less negative when stock-return volatility increases. This is because expected returns are particularly sensitive to return shocks when the predictive variable is far from its long-term mean, uncertainty is high, and stock-return volatility is low.

In the empirical section of the paper, we test the dependence of attention to news and the risky investment share on the state variables of our model. We first calibrate the model to S&P 500 returns and define the predictive variable as the S&P 500 earnings-to-price ratio. Using this dataset, we then build three *model-implied* time series: uncertainty, attention, and risky investment share. We first compare these time series to their empirical counterparts and find a strong positive correlation in each case. We also provide evidence that our model-implied measure of attention is counter-cyclical and strongly predicts the VIX index. Consistent with the predictions of the model, we find that the empirical proxy for attention to news is indeed U-shaped in the predictive variable, it increases with both uncertainty and the magnitude of the predictive coefficient, and it decreases with stock-return volatility. In addition, we show that the empirical proxy for the risky investment share increases with the Sharpe ratio of the stock and features a negative hedging demand whose shape mirrors the model's prediction. We therefore conclude that our rational setup accurately describes investors' dynamic asset and attention allocation behavior.

This paper complements a large body of literature that considers portfolio selection problems with stochastic expected returns, stochastic volatility, incomplete information, and uncertainty about predictability.[8] In the costly information acquisition literature, Detemple and Kihlstrom (1987) analyze the demand for information and the equilibrium price of information in the context of a production economy. Veldkamp (2006a,b) shows that costly information acquisition helps explain excess co-movement and the simultaneous increases in emerging markets' media coverage and equity prices. The most closely related paper is Huang and Liu (2007), who consider a dynamic

[8]For portfolio selection problems with stochastic expected returns see Kim and Omberg (1996), Brandt (1999), Campbell and Viceira (1999), Ait-Sahalia and Brandt (2001), Wachter (2002), Detemple, Garcia, and Rindisbacher (2003) with stochastic volatility see Chacko and Viceira (2005), Liu (2007); with incomplete information see Detemple (1986), Gennotte (1986), Dothan and Feldman (1986), Feldman (1989), David (1997), Brennan (1998), Veronesi (1999, 2000), Brennan and Xia (2001); and with uncertainty about predictability see Kandel and Stambaugh (1996), Barberis (2000), Xia (2001), and Brandt et al. (2005).

portfolio choice problem with static costly information choice. That is, the investor dynamically chooses her portfolio, but the accuracy and frequency of information is chosen at time zero only. The investor optimally acquires a flow of information that has limited accuracy and limited frequency, which potentially makes her under- or over-invest. The frequency and accuracy of information acquisition is shown to be decreasing and increasing with risk aversion, respectively.

Our paper contributes to this literature by considering a dynamic portfolio choice and information acquisition problem in the presence of uncertainty about stock-return predictability. Our study sheds light on the dynamic relation between attention, risky investments, and the relevant state variables. More precisely, we show that attention is U-shaped in the predictive variable, increasing with uncertainty about the predictive coefficient, decreasing with stock-return volatility, and increasing with the magnitude of the predictive coefficient. As a result, the relation between attention and the risky investment share is positive when expected returns are high and negative when expected returns are low. These predictions are first described in our theoretical framework, and are then shown to be supported by the data.

The remainder of the paper is organized as follows. Section 2 describes the economy and examines the optimal attention, consumption, and portfolio choice problem of the investor. Section 3 calibrates the parameters of the model and describes the results. Section 4 performs the empirical analysis. Section 5 concludes. The Appendix contains all proofs and computational details.

2 The model

Consider an economy populated by an investor with utility function defined by

$$U(c) = \mathbb{E}\left(\int_0^\infty e^{-\delta s} u(c_s) ds\right), \tag{1}$$

where c_t is the consumption at time t, δ is the subjective discount rate, and $u(c)$ is an increasing and concave function of c differentiable on $(0, \infty)$.

The investor continuously trades one risk-free asset paying a constant rate of return r^f, and one risky asset (the *stock*) whose price dynamics satisfy

$$\frac{dP_t}{P_t} = \mu_t dt + \sqrt{V_t} dB_{P,t}, \tag{2}$$

where μ_t is the instantaneous expected return on the stock and V_t is the instantaneous variance of stock returns.

The investor operates under partial knowledge of the economy. Specifically, the expected return μ_t is unobservable (Brennan, 1998; Xia, 2001; Ziegler, 2003), but the investor knows that it is an affine function of an observable state variable y_t. That is, the expected return satisfies

$$\mu_t = \bar{\mu} + \beta_t (y_t - \bar{y}), \tag{3}$$

where β_t is an *unobservable* predictive coefficient (Xia, 2001). The observable predictive variable y_t and the unobservable predictive coefficient β_t evolve according to the following diffusion processes:

$$dy_t = \lambda_y (\bar{y} - y_t) dt + \sigma_y dB_{y,t} \tag{4}$$

$$d\beta_t = \lambda_\beta (\bar{\beta} - \beta_t) dt + \sigma_\beta dB_{\beta,t}, \tag{5}$$

where we assume that $\bar{y}$, λ_y, σ_y, $\bar{\beta}$, λ_β, and σ_β are known constants. The variance of stock returns is observable and follows a square-root process (Heston, 1993; Liu, 2007):

$$dV_t = \lambda_V (\bar{V} - V_t) dt + \sigma_V \sqrt{V_t} dB_{V,t}, \tag{6}$$

where $\bar{V}$, λ_V and σ_V are known constants. The four Brownian motions $B_{P,t}$, $B_{y,t}$, $B_{\beta,t}$, and $B_{V,t}$ are independent from each other.

2.1 The inference process: active learning

Given the dynamics of the state variables described above, the investor's problem consists of inferring the predictive coefficient β_t before choosing an optimal portfolio and consumption rule that maximizes the expected lifetime utility of consumption.

The investor has the opportunity to actively learn about return predictability, i.e. to collect arbitrarily accurate information about the predictive coefficient β_t. This is achieved by acquiring a news signal with the following dynamics:

$$ds_t = \beta_t dt + \frac{1}{\sqrt{a_t}} dB_{s,t}, \tag{7}$$

where $B_{s,t}$ is an Brownian motion independent of $B_{P,t}$, $B_{y,t}$, $B_{\beta,t}$, and $B_{V,t}$.

The dynamics of the news signal (7) are interpreted as follows. Assume the investor acquires n_t signals of equal precision s_t^j, $j = 1, \ldots, n_t$ at time t:

$$ds_t^j = \beta_t dt + \sigma_s dB_t^j, \quad j = 1, \ldots, n_t \tag{8}$$

where B_t^j is independent of $B_{P,t}$, $B_{y,t}$, $B_{\beta,t}$, and $B_{V,t}$ for all j and $B_t^j \perp B_t^i$, $\forall j \neq i$. Aggregating yields the following dynamics of the aggregate signal s_t acquired by the investor

$$ds_t \equiv \frac{1}{n_t} \sum_{j=1}^{n_t} ds_t^j = \beta_t dt + \frac{\sigma_s}{\sqrt{n_t}} dB_{s,t}, \tag{9}$$

where $B_{s,t}$ is independent from $B_{P,t}$, $B_{y,t}$, $B_{\beta,t}$, $B_{V,t}$. Setting $\frac{\sigma_s}{\sqrt{n_t}} \equiv \frac{1}{\sqrt{a_t}}$ in Equation (9) leads to Equation (7). That is, the investor controls the accuracy a_t of the aggregate signal by choosing the number of individual signals n_t she acquires. When the investor is attentive to news, the number of individual signals she acquires is large and the aggregate signal is accurate. When the investor is inattentive to news, the number of individual signals she acquires is small and the aggregate signal is inaccurate. Given this, we call a_t the investor's *attention to news*.

Denote by $\mathcal{F}_t$ the information set of the investor at time t. This information set includes: realized returns defined in (2), changes in the predictive variable defined in (4), changes in the instantaneous variance of stock returns defined in (6), and changes in the signal defined in (7). This last source of information is the focus of our paper. The key feature is that the investor is able to change her information acquisition policy by controlling the magnitude of the noise in the signal (7) *at any point in time*. This results in a control problem with an endogenous information structure.

Let us denote by $\widehat{\beta}_t \equiv \mathbb{E}\left[\beta_t | \mathcal{F}_t\right]$ the estimated predictive coefficient and its posterior variance by $\nu_t \equiv \mathbb{E}[(\beta_t - \widehat{\beta}_t)^2|\mathcal{F}_t]$. Thus,

$$\beta_t \sim N(\widehat{\beta}_t, \nu_t), \tag{10}$$

where $N(m,v)$ denotes the Normal distribution with mean m and variance v. Henceforth, we refer to the estimated predictive coefficient $\widehat{\beta}_t$ as the *filter* and to the posterior variance ν_t as the *uncertainty*.

The dynamics of the state variables observed by the investor are obtained from standard filtering results (Liptser and Shiryaev, 2001) and are provided in Proposition 1 below.

Proposition 1. *The dynamics of the observed state variables satisfy*

$$\frac{dP_t}{P_t} = \left(\bar{\mu} + \widehat{\beta}_t(y_t - \bar{y})\right) dt + \begin{bmatrix} \sqrt{V_t} & 0 & 0 & 0 \end{bmatrix} d\widehat{B}_t^{\perp} \tag{11}$$

$$ds_t = \widehat{\beta}_t dt + \begin{bmatrix} 0 & 0 & 0 & \frac{1}{\sqrt{a_t}} \end{bmatrix} d\widehat{B}_t^{\perp} \tag{12}$$

$$dy_t = \lambda_y(\bar{y} - y_t)dt + \begin{bmatrix} 0 & \sigma_y & 0 & 0 \end{bmatrix} d\widehat{B}_t^{\perp} \tag{13}$$

$$dV_t = \lambda_V(\bar{V} - V_t)dt + \begin{bmatrix} 0 & 0 & \sigma_V\sqrt{V_t} & 0 \end{bmatrix} d\widehat{B}_t^{\perp}. \tag{14}$$

The dynamics of the filter and uncertainty are

$$d\widehat{\beta}_t = \lambda_\beta(\bar{\beta} - \widehat{\beta}_t)dt + \begin{bmatrix} \frac{\nu_t(y_t - \bar{y})}{\sqrt{V_t}} & 0 & 0 & \nu_t\sqrt{a_t} \end{bmatrix} d\widehat{B}_t^{\perp} \tag{15}$$

$$\frac{d\nu_t}{dt} = -\left(\frac{(y_t - \bar{y})^2}{V_t} + a_t\right)\nu_t^2 - 2\lambda_\beta\nu_t + \sigma_\beta^2, \tag{16}$$

where $\widehat{B}_t^{\perp} \equiv \begin{bmatrix} \widehat{B}_{1,t}^{\perp} & \widehat{B}_{2,t}^{\perp} & \widehat{B}_{3,t}^{\perp} & \widehat{B}_{4,t}^{\perp} \end{bmatrix}^{\top}$ *is a 4-dimensional vector of independent Brownian motions under the investor's observation filtration.*

Proof. See Liptser and Shiryaev (2001).

Equations (15) and (16) describe the investor's updating rule regarding the expectation and variance of the predictive coefficient. The instantaneous change in the filter is driven by four sources of information: realized returns, changes in the predictive variable, changes in volatility, and changes in the news signal. As Equation (15) shows, the investor assigns stochastic weights to these four sources of information.[9] As we will describe below, the size of these weights depend on the relative informativeness of each source of information.

Equation (16) describes the change in uncertainty when the investor controls her attention to news. Uncertainty is locally deterministic and decreases faster when the investor's attention is high. The decline in uncertainty is weaker when the predictive coefficient is more persistent (i.e. low λ_β) or when the volatility of realized returns V_t is high. Finally, the last term in Equation (16) shows that the larger the volatility of the predictive coefficient, the stronger the increase in uncertainty over time.

The informativeness of the signal depends on investor's attention, which impacts learning in two ways. First, it has a direct impact on the instantaneous volatility of the filter in Equation (15). Second, it drives the drift of uncertainty in Equation (16). We analyze each of these two effects separately. To facilitate our discussion, we refer to $d\widehat{B}^{\perp}_{1,t}$ as *return shocks* and to $d\widehat{B}^{\perp}_{4,t}$ as *news shocks*.

2.1.1 The impact of attention on the filter

The magnitude of the impact of return shocks and news shocks on the filter depend on the uncertainty ν_t, on the difference between the predictive variable and its long-term mean $y_t - \bar{y}$, and on investor's attention a_t. The following example provides insight on how the investor updates her beliefs using each piece of information.

Suppose that $y_t > \bar{y}$. Then, an unexpectedly high return ($d\widehat{B}^{\perp}_{1,t} > 0$) means that the current estimate of β_t is too low, and the investor adjusts $\widehat{\beta}_t$ upwards. The opposite happens when $y_t < \bar{y}$: An unexpectedly high return means that the current estimate of β_t is too high, and the investor adjusts $\widehat{\beta}_t$ downwards. Hence, the first coefficient in the diffusion of the filter has the same sign as $y_t - \bar{y}$ (see also Xia (2001) for a similar interpretation).

An additional component drives the filter through active learning from news shocks. When attention is high, the signal becomes more informative and therefore the investor increases the weight assigned to news shocks, as can be seen from the last coefficient in the diffusion of the filter.

[9] Because the Brownian motions $B_{P,t}$, $B_{y,t}$, $B_{\beta,t}$, $B_{V,t}$, and $B_{s,t}$ are uncorrelated, shocks to the predictive variable y_t and to return variance V_t do not impact the investor's estimate of β_t. Hence, the second and third components of the diffusion of $\widehat{\beta}_t$ are both equal to zero.

Overall, the instantaneous variance of the filter is an increasing function of attention:

$$\text{Var}[d\widehat{\beta}_t] = \nu_t^2 \left(\frac{(y_t - \bar{y})^2}{V_t} + a_t \right). \tag{17}$$

As attention converges to infinity, the instantaneous variance of the filter converges to its upper bound σ_β^2. This upper bound represents the variance of the filter when the predictive coefficient β_t is perfectly observable.

2.1.2 The impact of attention on uncertainty

The predictive variable y_t is a key driver of the dynamics of uncertainty. Intuitively, if y_t is close to its long-term mean, learning from realized returns becomes ineffective in estimating β_t because the signal-to-noise ratio is very low. Therefore, the reduction in uncertainty is weak when $y_t - \bar{y} \approx 0$. In contrast, when y_t is far away from its long-term mean, realized returns offer valuable information on the predictive coefficient β_t and uncertainty decreases faster.

Equation (16) shows that uncertainty decreases faster when attention is high. It is worth noting that uncertainty does not converge to a "steady state" in this model because three stochastic variables, namely the predictive variable y_t, the volatility of asset returns V_t, and investor's attention a_t, drive its dynamics.

2.2 Properties of the estimated expected returns

The following lemma describes the properties of the *estimated expected return* of the stock, defined as:

$$\widehat{\mu}_t = \bar{\mu} + \widehat{\beta}_t (y_t - \bar{y}). \tag{18}$$

Lemma 1. *The dynamics of the estimated expected return follow*

$$d\widehat{\mu}_t = (\lambda_y + \lambda_\beta) \left(\bar{\mu} + \frac{\bar{\beta}\lambda_\beta (y_t - \bar{y})}{\lambda_y + \lambda_\beta} - \widehat{\mu}_t \right) dt + \begin{bmatrix} \frac{(y_t - \bar{y})^2 \nu_t}{\sqrt{V_t}} & \sigma_y \widehat{\beta}_t & 0 & \nu_t \sqrt{a_t} (y_t - \bar{y}) \end{bmatrix} d\widehat{B}_t^\perp. \tag{19}$$

The mean square error of this estimate (i.e. the uncertainty about expected returns, which we denote *hereafter by η_t) satisfies*

$$\eta_t \equiv \mathbb{E}\left[(\mu_t - \widehat{\mu}_t)^2 \middle| \mathcal{F}_t\right] = (y_t - \bar{y})^2 \nu_t. \tag{20}$$

The instantaneous variance of the estimated expected return is

$$\mathrm{Var}[d\widehat{\mu}_t] = \nu_t^2 (y_t - \bar{y})^2 \left(a_t + \frac{(y_t - \bar{y})^2}{V_t}\right) + \sigma_y^2 \widehat{\beta}_t^2. \tag{21}$$

$\mathrm{Var}[d\widehat{\mu}_t]$ *is a monotone increasing function of attention. Its maximum depends on both $\widehat{\beta}_t$ and y_t and is given by*

$$\lim_{a_t \to \infty} \mathrm{Var}[d\widehat{\mu}_t] = \sigma_\beta^2 (y_t - \bar{y})^2 + \sigma_y^2 \widehat{\beta}_t^2. \tag{22}$$

Expected returns mean-revert at speed $\lambda_y + \lambda_\beta$ to a stochastic level that depends on the predictive variable. If the long-term mean $\bar{\beta}$ is assumed to be zero—which means that there is no predictability on average—then the stochastic level simplifies to the constant $\bar{\mu}$.

As shown in Equation (19), when the filter $\widehat{\beta}_t$ is large, expected returns react to changes in the predictive variable y_t (the second component of the diffusion). Furthermore, if investor's attention is high, expected returns react to news shocks, but *only* when $y_t \neq \bar{y}$ (the fourth component of the diffusion). This concurs with the relation between returns and the predictive variable described in Equation (3), whereby more news on the predictive coefficient β_t—no matter how accurate—is not going to change investor's view about expected returns if $y_t - \bar{y} = 0$.

High uncertainty ν_t magnifies the sensitivity of expected returns to return shocks (the first component of the diffusion). When $y_t - \bar{y} \neq 0$, an increase in uncertainty increases the variance of expected returns. This is shown in Equation (21). Furthermore, higher attention (or more accurate news) increases the variance of expected returns by making them more sensitive to news shocks. The variance of expected returns reaches its maximum when attention converges to infinity, as shown in Equation (22).[10]

To summarize, attention to news drives two important factors, the variance of expected returns

[10] We derive the latter equation by applying Itô's lemma to the expected return in Equation (3) and by assuming that the predictive coefficient β_t is perfectly observable.

and the drift of uncertainty. More attention increases the sensitivity of expected returns to news shocks and therefore augments their variance. At the same time, more attention yields lower future uncertainty by magnifying the negative component of its drift.

2.3 Optimal attention, portfolio choice, and consumption

Turning now to the investor's optimization problem, we consider $\widehat{\mu}_t$ as a state variable instead of $\widehat{\beta}_t$, with the aim to better interpret and characterize our results. The investor's problem is to choose consumption c_t, attention to news a_t, and the risky investment share w_t so as to maximize her expected lifetime utility of consumption conditional on her information set at time t, $\mathcal{F}_t$. That is, the investor's maximization problem is written

$$J(W_t, \widehat{\mu}_t, y_t, V_t, \nu_t) \equiv \max_{c,a,w} \mathbb{E}\left[\int_t^\infty e^{-\delta(s-t)} u(c_s) ds \middle| \mathcal{F}_t\right], \tag{23}$$

subject to the budget constraint

$$dW_t = \left[r^f W_t + w_t W_t \left(\widehat{\mu}_t - r^f\right) - c_t - \overline{K}_t\right] dt + w_t W_t \begin{bmatrix} \sqrt{V_t} & 0 & 0 & 0 \end{bmatrix} d\widehat{B}_t^{\perp}. \tag{24}$$

We assume that the total information cost, $\overline{K}_t$, is linear in wealth:

$$\overline{K}_t = K(a_t) W_t. \tag{25}$$

This assumption reflects the fact that, similar to the price of financial securities, the price of information tends to increase as time passes (i.e., the price of information features an exponential time trend). Furthermore, the above cost function preserves the homogeneity of the value function in wealth and therefore implies that attention, the consumption-to-wealth ratio, and the information cost-to-wealth ratio are independent of wealth at the optimum. Finally, the per-unit of wealth cost function, $K(a_t)$, is assumed to be increasing and convex in attention, which implies that perfect information ($a_t \to \infty$) cannot be attained, and thus the investor is never able to observe the true level of expected returns.

Since attention does not depend on wealth at the optimum, it is a function of the expected return $\widehat{\mu}_t$, the predictive variable y_t, stock-return variance V_t, and uncertainty ν_t. These state variables do not feature any trend and thus neither does attention. The stationary dynamics of optimal attention implied by the cost function in (25) are therefore consistent with Sims (2003), who argues that investors have limited information-processing capacity, i.e. investors' attention is bounded. If instead attention were increasing with wealth, then it would tend to increase as time passes and would therefore not be bounded.

Two additional considerations about the information cost function (25) are worth mentioning. First, this specification implies that it is more costly to acquire information when one gets wealthier. In reality, it could be the case that the cost is actually decreasing in wealth, because a wealthier investor may acquire more information and thus may get a volume discount on the cost. In Appendix C, we discuss the implications of a more general information cost specification, which allows $\overline{K}_t$ to be *ex ante* either increasing with wealth, independent of wealth, or decreasing with wealth. We show that such a specification does not change the qualitative results of our paper.

Second, the specification (25) implies that if two investors (A and B) start with different levels of wealth W_0^A and W_0^B, then they will incur different costs of information at any given time t (because the wealth at time t will be different across investors). To avoid such situations in which there is a discount or an overcharge across investors, the cost function can be normalized by dividing by the initial level of wealth. This modification clearly does not alter our results, but ensures that poor and rich investors pay the same price for the same piece of information at any given time t.

To summarize, given the total information cost (25), if the investor chooses to be inattentive to news—and therefore to learn using only the information provided by the price P_t, the predictive variable y_t, and the variance of stock returns V_t—then her entire wealth is invested in the financial market. In contrast, if the investor decides to pay attention to news ($a_t > 0$), then a positive fraction of her wealth flows to the information market in order to pay for research expenditures. Attention, therefore, can be perceived as a non-financial security in the investor's portfolio.

Proposition 2. *The optimal consumption c_t^*, risky investment share w_t^*, and attention to news a_t^**

are given by

$$c_t^* = u_c^{-1}(J_W) \tag{26}$$

$$w_t^* = \frac{\widehat{\mu}_t - r^f}{V_t} \frac{-J_W}{J_{WW} W_t} + \frac{\nu_t (y_t - \bar{y})^2}{V_t} \frac{-J_{W\mu}}{J_{WW} W_t} \tag{27}$$

$$a_t^* = \Phi \left(\frac{1}{2 J_W W_t} \left(\nu_t^2 (y_t - \bar{y})^2 J_{\mu\mu} - 2\nu_t^2 J_\nu \right) \right), \tag{28}$$

where $\Phi(\cdot)$ *is a positive and increasing function defined as the inverse of the derivative of the cost function,* $\Phi(\cdot) \equiv K'(\cdot)^{-1}$.

Proof. See Appendix A.

Equation (26) is the standard optimal consumption derived by Merton (1971). The optimal risky investment share, expressed in Equation (27), comprises a myopic and a hedging portfolio (Merton, 1971). The hedging term represents the effect of parameter learning and significantly impacts the asset allocation decision (Brennan, 1998; Xia, 2001). It is positive if $\gamma < 1$ and negative if $\gamma > 1$. It vanishes when the state variables are observable (i.e. when $\nu_t = 0$) because, by assumption, none of these variables are correlated to returns.

Our object of focus is the optimal attention a_t^*, expressed in Equation (28). Since the function $\Phi(\cdot)$ is positive and increasing, we can directly analyze the term in brackets. Two factors drive the optimal level of attention: the *state risk aversion factor* $J_{\mu\mu}$, which measures the extent to which the investor (dis)likes variations in expected returns and the *uncertainty factor* J_ν, which measures the extent to which the investor (dis)likes uncertainty. Recall from Section 2.2 that attention drives both the variance of expected returns and the drift of uncertainty. The state risk aversion factor $J_{\mu\mu}$ is multiplied by $\nu_t^2(y_t - \bar{y})^2$, which is the marginal effect of attention on the variance of expected returns (see Equation (21)). The uncertainty factor J_ν is multiplied by $-\nu_t^2$, which is the marginal effect of attention on the drift of ν_t (see Equation (16)).

Because our setup features mean-reverting expected returns, the value function is convex in $\widehat{\mu}_t$, which yields $J_{\mu\mu} > 0$ (Kim and Omberg, 1996). That is, the investor prefers expected return volatility because it creates the possibility of trends that she can exploit. The higher the volatility, the higher the convexity of the value function and thus the investor pays greater attention to accurately estimate the predictive coefficient and efficiently exploit future trends.

Two opposing forces determine the sign of the uncertainty factor J_ν. First, the investor dislikes uncertainty; second, the investor likes uncertainty because it leads to higher expected return volatility and a higher likelihood of trends. Depending on which of these two forces dominates, the uncertainty factor J_ν is positive or negative. If it is negative, the investor acquires more information to reduce uncertainty. If it is positive, the investor acquires less information in order to keep expected return volatility high and take advantage of trends.

Overall, Proposition 2 implies that the investor optimally chooses to be more attentive to news when (i) uncertainty is high and (ii) the predictive variable moves away from its long-term mean. These results are independent on the investor's utility function. As Equation (28) shows, the effects of uncertainty and the predictive variable on attention reinforce each other, yielding high attention in environments characterized by high uncertainty and by a large difference between the predictive variable and its long-term mean.

2.4 CRRA utility and quadratic attention cost

To illustrate the effects of optimal learning about predictability, we assume that the investor has a CRRA utility function with risk aversion parameter γ. In addition, the per-unit of wealth information cost function is specified in quadratic form:

$$K(a_t) = k a_t^2, \tag{29}$$

where $k > 0$ is the information cost parameter. In this case, the inverse of the derivative of the function $K(.)$ satisfies

$$\Phi(x) = \frac{x}{2k}. \tag{30}$$

Under these assumptions, the value function J can be written as

$$J(W_t, \widehat{\mu}_t, y_t, V_t, \nu_t) = \frac{W_t^{1-\gamma}}{1-\gamma}\phi(\widehat{\mu}_t, y_t, V_t, \nu_t). \tag{31}$$

Computing the partial derivatives of J as a function of the partial derivatives of ϕ and substituting them into the first-order conditions yields the optimal consumption, risky investment share, and attention to news provided in Proposition 3 below.

Proposition 3. *With CRRA utility and quadratic attention cost, the optimal consumption* c_t^**, risky investment share* w_t^**, and attention to news* a_t^* *are given by*

$$c_t^* = \phi^{-1/\gamma} W_t \tag{32}$$

$$w_t^* = \frac{\widehat{\mu}_t - r^f}{\gamma V_t} + \frac{\nu_t (y_t - \bar{y})^2}{\gamma V_t} \frac{\phi_\mu}{\phi} \tag{33}$$

$$a_t^* = \nu_t^2 \left(\frac{\phi_\nu}{\phi} \frac{1}{2k(\gamma - 1)} + \frac{-\phi_{\mu\mu}}{\phi} \frac{(y_t - \bar{y})^2}{4k(\gamma - 1)} \right). \tag{34}$$

The optimal consumption defined in Equation (32) is well-known (Merton, 1971) and does not require any further analysis. The optimal risky investment share defined in Equation (33) is analyzed by Xia (2001) in a setup with constant stock-return volatility.[11] We discuss the dependence of the risky investment share and its hedging components on the state variables in Section 3.3.

The optimal attention defined in Equation (34) becomes a strictly increasing quadratic function of uncertainty. We discuss the dependence of the investor's attention on the state variables in Section 3.2.

Although an indirect dependence of attention on the variance of stock returns, V_t, arises through the function ϕ, the variance of stock return has no direct impact on the investor's attention. In Section 3.2 we show that the indirect impact of the return variance on attention is quantitatively weak, as opposed to the impact of uncertainty and the predictive variable.

3 Numerical results

In this section, we investigate the determinants of optimal attention and risky investment share. We first calibrate the parameters of the model. Then, we show that attention is a U-shaped function of the predictive variable, a decreasing function of the stock-return variance, and an increasing function of both the absolute value of the predictive coefficient and uncertainty. The risky investment share increases with the Sharpe ratio of the stock and features a negative hedging demand that is hump-

[11] In Xia (2001), the risky investment share features additional terms resulting from the correlations between realized returns and the predictive variable y_t, and between returns and the predictive coefficient β_t. These correlations are set to zero in the present case. Note also that, although a similar decomposition appears in Xia (2001), the attention level affects the shape of the value function, causing differences between the portfolio holdings obtained in the two models.

shaped in the predictive variable. Furthermore, the size of the hedging demand increases with uncertainty and decreases with stock-return volatility.

3.1 Calibration

We calibrate the parameters of the model using three datasets. First, we consider the S&P 500 earnings-to-price ratio to be the predictive variable y_t of S&P 500 returns r_t.[12] This dataset is at monthly frequency from 01/1950 to 12/2014, and is obtained from Robert Shiller's website.[13] The dynamics of the predictive variable provided in Proposition 1 imply that the long-term mean of y_t is $\bar{y}$, the long-term variance of y_t is $\sigma_y^2/(2\lambda_y)$, and $\text{cov}(y_{t+\Delta}, y_t)/\text{var}(y_t) = e^{-\lambda_y \Delta}$, where $\Delta = 1/12 = 1$ month. Solving these three moment conditions yields the parameters $\bar{y}$, σ_y, and λ_y.

Second, we jointly estimate the long-term expected return $\bar{\mu}$ and the time-varying predictive coefficients $\widehat{\beta}_t$ by performing multivariate 60-month rolling window regressions of 1-month-ahead returns on current demeaned earnings-to-price ratios. That is, $\bar{\mu}$ and $\widehat{\beta}_t$ satisfy

$$r_{t+\Delta} = \left[\bar{\mu} + \widehat{\beta}_{j\Delta+60\Delta}(y_t - \bar{y})\right]\Delta + \epsilon_{t+\Delta}, \quad t \in (j\Delta, j\Delta + 59\Delta), \tag{35}$$

where $j = 0, \ldots, N-1$ is the window's index, N is the total number of windows, and ϵ_t is a random variable with mean 0 and variance $V_t\Delta$. The dynamics of the predictive coefficient provided in Proposition 1 imply that the long-term mean of $\widehat{\beta}_t$ is $\bar{\beta}$, the long-term variance of $\widehat{\beta}_t$ can be approximated by $\sigma_\beta^2/(2\lambda_\beta)$, and $\text{cov}(\widehat{\beta}_{t+\Delta}, \widehat{\beta}_t)/\text{var}(\widehat{\beta}_t) = e^{-\lambda_\beta \Delta}$. Solving these three moment conditions yields the parameters $\bar{\beta}$, σ_β, and λ_β.

Third, we compute the demeaned returns ϵ_t as follows

$$\epsilon_{j\Delta+61\Delta} = r_{j\Delta+61\Delta} - \left[\bar{\mu} + \widehat{\beta}_{j\Delta+60\Delta}(y_{j\Delta+60\Delta} - \bar{y})\right]\Delta, \tag{36}$$

and estimate their conditional variance $V_t\Delta$ by fitting a GARCH(1,1) model (Bollerslev, 1986). The dynamics of the stock return variance provided in Proposition 1 imply that the long-term mean of V_t is $\bar{V}$, the long-term variance of V_t is $\sigma_V^2\bar{V}/2\lambda_V$, and $\text{cov}(V_{t+\Delta}, V_t)/\text{var}(V_t) = e^{-\lambda_V \Delta}$. Solving

[12] Several other predictive variables have been identified. They include past market returns, the dividend yield, nominal interest rates, and expected inflation among others. See Goyal and Welch (2008) for a comprehensive survey.

[13] http://www.econ.yale.edu/ shiller/data.htm

these three moment conditions yields the parameters $\bar{V}$, σ_V, and λ_V.

The estimated parameter values are provided in Table 1. In addition, we set the risk-free rate to its historical mean $r^f = 5.08\%$. The cost parameter is $k = 0.1$, risk aversion is $\gamma = 3$, and the subjective discount rate is $\delta = 0.01$.

It is worth noting that the parameters λ_β, $\bar{\beta}$, and σ_β imply that the predictive coefficient β_t can be both positive and negative.[14] The predictive regression literature (e.g. Cochrane, 2008) finds a relatively weak positive unconditional relation between future returns and the earnings-to-price ratio with annual data; with monthly data (as in this paper), unconditional regressions are not statistically significant at conventional levels (Stambaugh, 1999).[15] In our *conditional* regressions, the 60-month relationship between the earnings-to-price ratio and future returns shows both a positive and negative direction of predictability, which may appear difficult to rationalize from an economic point of view.

Although the objective of this paper is not to provide an economic story which rationalizes the switching sign of the relationship between the earnings-to-price ratio and future returns, we briefly discuss a plausible mechanism here. Consider an economy with time-varying and mean-reverting macroeconomic uncertainty, populated by a representative agent with CRRA preferences and time-varying risk aversion. Let us assume that the price dynamics in Equation (2) are determined in general equilibrium by the representative agent. Furthermore, suppose that the risk aversion of the representative agent is countercyclical, i.e., it is higher when macroeconomic uncertainty is high.[16]

In equilibrium, prices are affected by the income and substitution effects. Because of the countercyclical risk aversion assumption, the income effect dominates during periods of high uncertainty (i.e., the risk aversion of the representative agent is above one). In this case, the consumption smoothing motive is strong and high uncertainty leads the investor to consume less and therefore to save more (i.e. invest more in the stock), which implies high prices i.e. low earnings-to-price ratios. Conversely, the substitution effect dominates during periods of low uncertainty (i.e., the

[14]The results are similar if we use the dividend-price ratio instead of the earnings-to-price ratio, or if we use the inverse of the cyclically adjusted price-earnings ratio (CAPE).

[15]In our case, the unconditional relation is positive, but statistically insignificant.

[16]Fluctuations in economic uncertainty are well-documented (Bloom, 2009; Jurado, Ludvigson, and Ng, 2015). Theoretical models with time-varying uncertainty include Veronesi (1999, 2000) and Andrei, Carlin, and Hasler (2017). Time-varying risk aversion arises when investors derive utility form consumption relative to a habit (Campbell and Cochrane, 1999), or when utility depends on recent investment performance relative to some historical benchmark (Barberis, Huang, and Santos, 2001). Evidence for countercyclical risk aversion is provided in Cohn, Engelmann, Fehr, and Maréchal (2015).

Parameter	Symbol	Value
Mean-reversion speed of stock-return variance	λ_V	1.9592
Long-term mean of stock-return variance	$\bar{V}$	0.0158
Volatility of stock-return variance	σ_V	0.1760
Mean-reversion speed of earning-to-price ratio	λ_y	0.1163
Long-term mean of earning-to-price ratio	$\bar{y}$	0.0686
Volatility of earning-to-price ratio	σ_y	0.0136
Long-term expected return	$\bar{\mu}$	0.0685
Mean-reversion speed of β	λ_β	0.2151
Long-term mean of β	$\bar{\beta}$	0.0021
Volatility of β	σ_β	0.9531

Table 1: Calibrated Parameter Values.

risk aversion is below one), and thus the investor is willing to forego present consumption to have more in the future. Low uncertainty about future consumption leads the investor to consume less and invest more in the stock, which again implies high prices i.e. low earnings-to-price ratios. In this—admittedly stylized, but not unrealistic—example, when uncertainty mean-reverts, there is a negative relation between the earnings-to-price ratio and future returns if the risk aversion switches from above one (when uncertainty is high) to below one (when uncertainty is low). Conversely, the relationship between the earnings-to-price ratio and future returns becomes positive in periods where risk aversion remains either between zero and one or above one.[17]

Using the parameter values reported in Table 1, we numerically solve the partial differential equation resulting from specification (31) by applying the Chebyshev collocation method (Judd, 1998). Further details are provided in Appendix B, where we specify the two boundary conditions, we discuss the transversality condition (Merton, 1998), and finally we measure the accuracy of the solution algorithm (which is of order 10^{-29}).

3.2 Optimal attention

Four state variables impact the optimal attention: the predictive variable y_t, the uncertainty ν_t, the stock-return variance V_t, and the predictive coefficient $\widehat{\beta}_t$. Since the predictive variable is the main determinant of expected returns, we choose to plot the optimal attention against the predictive

[17]The above example relies on significant fluctuations of the elasticity of intertemporal substitution (EIS). The value of this parameter is subject of ongoing debate (Epstein, Farhi, and Strzalecki, 2014). Quantitative assessments of the EIS come from experimental evidence (Brown and Kim, 2013) or from field experiments on health outcomes Thornton (2008). There are also several attempts to use actual data to estimate the EIS, and results differ widely. Hall (1988) finds a value between zero and 0.2, whereas Hansen and Singleton (1983) estimate values between 0.5 and two; it is not uncommon to find values as high as 10 (Eichenbaum, Hansen, and Singleton, 1988).

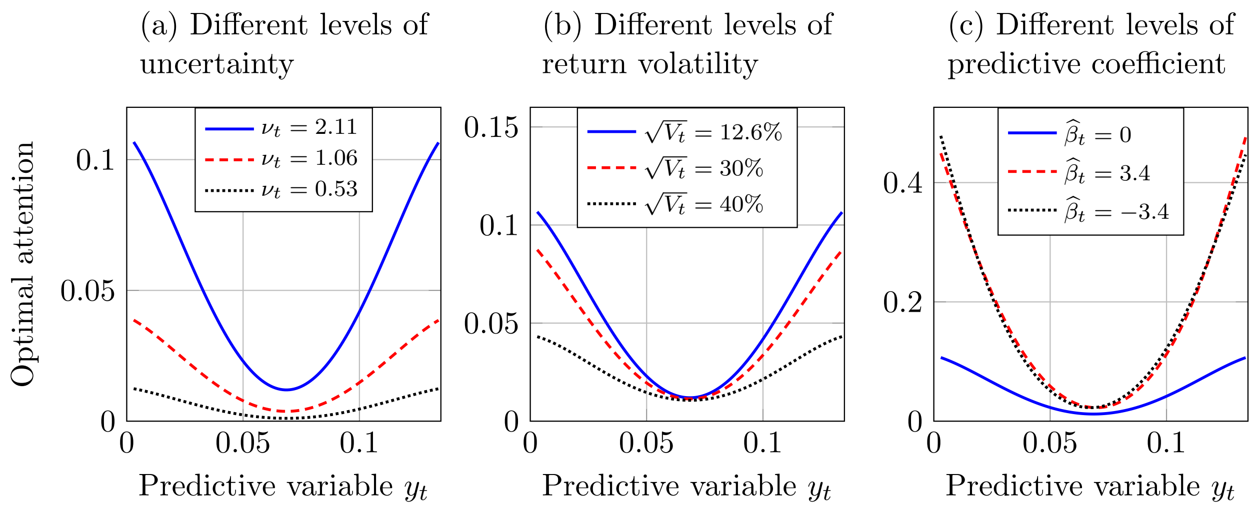


Figure 1: Impact of the predictive variable on attention.
The three panels depict the relation between attention and the predictive variable. We plot three curves corresponding to three different levels of uncertainty in panel (a), return volatility in panel (b), and the predictive coefficient in panel (c). If not stated otherwise, the state variables are $\widehat{\beta}_t = \bar{\beta} = 0$, $\sqrt{V_t} = \sqrt{\bar{V}} = 12.6\%$, and $\nu_t = \bar{\nu} = 2.11$. Parameter values are provided in Table 1.

variable for different values of the other state variables.

Figure 1 reports plots for three different levels of uncertainty in panel (a), return volatility in panel (b), and the estimated predictive coefficient in panel (c). The benchmark solid blue line is obtained by setting ν_t, V_t, and $\widehat{\beta}_t$ to their long-term levels $\bar{\nu}$, $\sqrt{\bar{V}}$, and $\bar{\beta}$.[18]

All three panels of Figure 1 confirm the intuition provided in Equation (34) that attention is a U-shaped function of the demeaned predictive variable $y_t - \bar{y}$. This is because when the predictive variable is close to its long-term mean, the investor knows that the expected return is equal to $\bar{\mu}$ and therefore has weak incentives to pay for information. In contrast, when the predictive variable is far from its long-term mean, there is a trend in expected returns that the investor can efficiently exploit, but only if the predictive coefficient is accurately estimated. The investor's optimal reaction to this situation is to pay attention, efficiently exploit the trend, and profit from it.

Panel (a) of Figure 1 shows that uncertainty drives investor attention in two ways. First, it increases the curvature of the U-shaped relation between attention and the predictive variable through the presence of ν_t^2 in Equation (34). Second, it slightly increases the level of the U-shaped

[18]The long-term uncertainty $\bar{\nu}$ is determined by solving $d\bar{\nu}/\bar{\nu} = 0$ conditional on setting $y_t = \bar{y}$, $V_t = \bar{V}$, and $a_t = 0$ in the dynamics of ν_t, which yields $\bar{\nu} = \sigma_\beta^2/2\lambda_\beta$. This is an upper bound of uncertainty because all the sources of information (i.e. the stock return, the predictive variable, the return variance, and the signal) are uninformative when $y_t = \bar{y}$ and $a_t = 0$.

relation between attention and the predictive variable. That is, higher uncertainty leads to higher attention. This is because when uncertainty is close to zero, the investor observes the predictive coefficient and feels no incentive to pay attention and learn about it. The greater the uncertainty, the less accurate the investors' estimates of the expected return, and therefore the larger their incentive to pay attention to news.

Panel (b) of Figure 1 shows that an increase in stock-return volatility decreases the curvature of the U-shaped relation between attention and the predictive variable. This is because an increase in the return volatility generates less informative returns, which decreases the volatility of expected returns (see Equation (21)). Because in this case trends are more difficult to detect, the convexity of the value function in expected returns is less pronounced. Since the convexity of the value function determines the curvature of the U-shaped relation between attention and the predictive variable (see Equation (34)), more return volatility leads to a weaker curvature and therefore to lower attention.[19]

Panel (c) of Figure 1 shows that an increase in the absolute value of the predictive coefficient increases the curvature of the U-shaped relation between attention and the predictive variable. This is because large positive or negative values of the predictive coefficient imply a high expected return volatility (see Equation (21)). Higher expected return volatility implies more opportunities to exploit trends, and thus a highly convex value function. This translates into a steeply curved U-shaped relation between attention and the predictive variable, and therefore into greater attention to news.

We now turn to the relation between attention and risk aversion, which is depicted in Figure 2. Consistent with Equation (34), an increase in risk aversion scales down the level of attention. This decreasing relation between attention and risk aversion comes from the fact that an increase in risk aversion decreases the investor's risky investment share (see the myopic component in Equation (33)). The smaller the risky investment share, the lower the investor's incentive to pay attention to news.

[19] This prediction is similar to the "ostrich effect," documented by Galai and Sade (2006), Karlsson, Loewenstein, and Seppi (2009) and Sicherman et al. (2015). The "ostrich effect" states that investors prefer to pay attention to their portfolios following positive news and "put their heads in the sand" when they expect to see bad news. Andries and Haddad (2014) provide an alternative explanation. In their model, investors are "disappointment averse" and thus are less attentive in riskier environments. In our case, riskier returns offer less marginal benefit for being attentive because the expected return becomes less responsive to information. Note that the "ostrich effect" commonly refers to *attention to wealth* (Abel, Eberly, and Panageas, 2007, 2013), whereas here we model investors' *attention to news*.

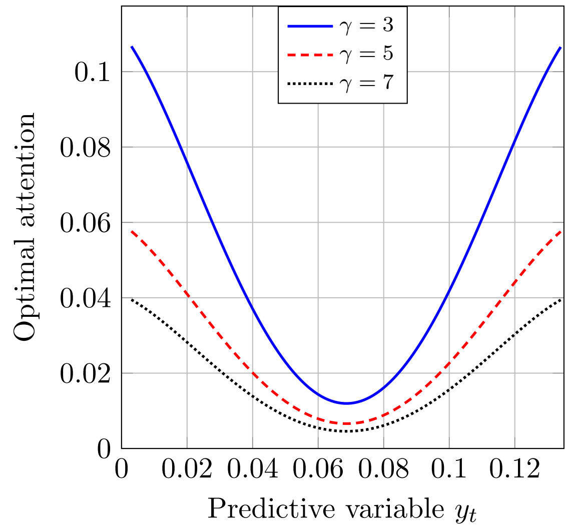


Figure 2: Impact of risk aversion on attention.
The figure depicts the relation between attention and the predictive variable for three different levels of risk aversion. State variables are $\widehat{\beta}_t = \bar{\beta} = 0$, $\sqrt{V_t} = \sqrt{\bar{V}} = 12.6\%$, and $\nu_t = \bar{\nu} = 2.11$. Parameter values are provided in Table 1.

Finally, we investigate the robustness of our results by performing a sensitivity analysis of the optimal attention with respect to changes in the dynamics of the predictive variable. More precisely, we analyze how attention responds to a change in the persistence and the volatility of the predictive variable. Panels (a) and (b) of Figure 3 show that attention increases when both the persistence and the volatility of the predictive variable decrease. This is because the persistence and the volatility of the predictive variable determine the conditional volatility of its future values. When either the persistence or the volatility is low, the conditional volatility of future values of the predictive variable is low.[20] Since the investor can efficiently exploit this "smoothness" to accurately predict future returns only if her estimate of the predictive coefficient is accurate, her optimal reaction to this situation is to pay more attention to news.

3.3 Optimal risky investment share

Figure 4 plots the optimal risky investment share against the predictive variable for different values of uncertainty in panel (a), stock-return volatility in panel (b), and the estimated predictive coefficient in panel (c). The benchmark solid blue line is obtained by setting ν_t, V_t, and $\widehat{\beta}_t$ to their

[20] The conditional variance of future values of the predictive variable satisfies: $\mathrm{Var}_t(y_s) = \left(1 - e^{-2\lambda_y(s-t)}\right)\sigma_y^2/2\lambda_y$. This function is increasing in σ_y and decreasing in λ_y.

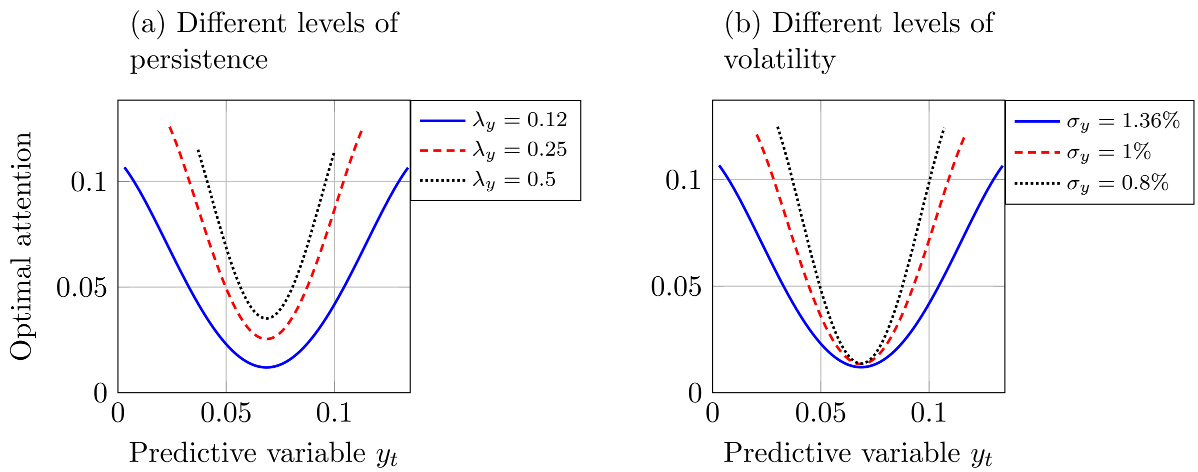


Figure 3: Impact of the dynamics of the predictive variable on attention.
Panels (a) and (b) depict the relation between attention and the predictive variable for three different levels of persistence of the predictive variable and its volatility, respectively. State variables are $\widehat{\beta}_t = \bar{\beta} = 0$, $\sqrt{V_t} = \sqrt{\bar{V}} = 12.6\%$, and $\nu_t = \bar{\nu} = 2.11$. If not stated otherwise, parameter values are provided in Table 1.

long-term levels $\bar{\nu}$, $\sqrt{\bar{V}}$, and $\bar{\beta}$.

Panel (a) of Figure 4 shows that the risky investment share is a hump-shaped function of the predictive variable and a decreasing function of uncertainty. This effect is driven by the hedging demand

$$H_t^* \equiv w_t^* - \frac{\widehat{\mu}_t - r^f}{\gamma V_t} = \frac{\nu_t (y_t - \bar{y})^2}{\gamma V_t} \frac{\phi_\mu}{\phi} < 0, \tag{37}$$

which reflects the investor's willingness to hedge against variations in expected returns.[21] Since returns and expected returns co-move positively (see Equations (11) and (19)), low returns imply low expected returns which triggers a negative hedging demand. Furthermore, the larger the expected return's loading on return shocks, the more negative the hedging demand is. According to Equation (19), this loading increases with both the predictive variable's deviation from its mean and uncertainty. That is, the hedging demand becomes more negative when both the predictive variable's deviation from its mean and uncertainty increase. This is confirmed by both Equation (37) and panel (a) of Figure 4. In addition, Equation (37) shows that the hedging demand becomes less negative when the stock-return volatility increases because the expected return's loading on return shocks is decreasing with the stock-return volatility.

[21] The hedging demand is negative because $J_\mu > 0$ and $J < 0$ imply that $\phi_\mu < 0$ and $\phi > 0$.

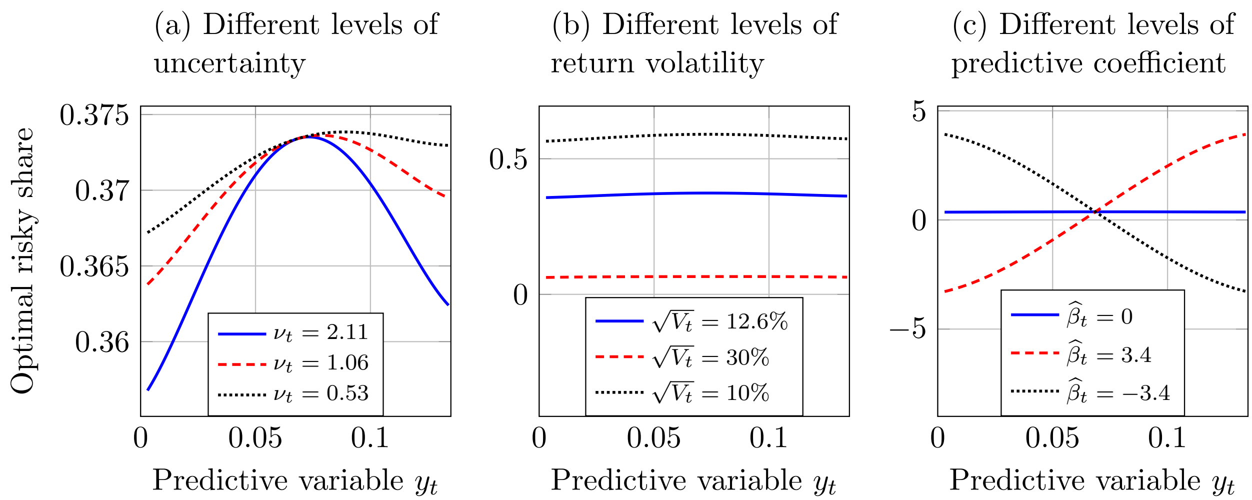


Figure 4: Impact of the predictive variable on the risky investment share.
The relation between the risky investment share and the predictive variable. We plot three curves corresponding to three different levels of uncertainty in panel (a), return volatility in panel (b), and the predictive coefficient in panel (c). If not stated otherwise, state variables are $\widehat{\beta}_t = \bar{\beta} = 0$, $\sqrt{V_t} = \sqrt{\bar{V}} = 12.6\%$, and $\nu_t = \bar{\nu} = 2.11$. Parameter values are provided in Table 1.

Panel (b) of Figure 4 shows that the risky investment share decreases with stock-return volatility (see myopic component in Equation (33)). Furthermore, the risky investment share increases with expected returns, which depend on the product of the predictive coefficient and the demeaned predictive variable. That is, the risky investment share increases with the predictive coefficient when the predictive variable is large and decreases with it when the predictive variable is small. As panel (c) of Figure 4 shows, the risky investment share is large whenever the product $\widehat{\beta}_t(y_t - \bar{y})$ is positive and large.

3.4 Cost of ignoring news

To quantify the benefits associated with paying attention to news, we compute the wealth certainty equivalent of the optimal strategy relative to that obtained when ignoring news (Xia, 2001; Das and Uppal, 2004; Liu, Peleg, and Subrahmanyam, 2010). That is, the cost of ignoring news is defined as the additional fraction of wealth required by an investor who ignores the news signal—equivalently, an investor who faces an infinite information cost—to reach the expected utility of an investor who optimally pays attention to news.

Table 2 reports the cost of ignoring news for different values of risk aversion γ and information cost parameter k. The cost of ignoring news decreases with both risk aversion and the information

Cost parameter $k \rightarrow$ **Risk aversion** $\gamma \downarrow$	0.01	0.05	0.1
3	512.46	31.4	19.14
5	54.46	9.49	5.24
7	37.38	4.49	2.42

Table 2: Cost of ignoring news (in bps).

The cost of ignoring news represents the additional fraction of wealth required by an investor who ignores the news signal to reach the expected utility of an investor who optimally pays attention to news. State variables are $\widehat{\beta}_t = \bar{\beta} = 0$, $y_t = \bar{y} = 6.9\%$ $\sqrt{V_t} = \sqrt{\bar{V}} = 12.6\%$, and $\nu_t = \bar{\nu} = 2.11$. Parameter values are provided in Table 1.

cost parameter. A high risk aversion implies a small share invested in the stock and therefore a weak incentive to pay attention to learn about the stock's expected return (see Figure 2). As a result, the optimal attention allocation strategy does not significantly differ from that of ignoring the news signal, i.e., the cost of ignoring information is small. Furthermore, the lower the cost parameter, the higher the optimal attention paid to news, and therefore the larger the cost of ignoring news. The cost of ignoring news can be significant, reaching as much as 5.1% of wealth when risk aversion and the information cost parameter are equal to 3 and 0.01, respectively.

4 Empirical analysis

In this section, we first show that there exists a positive and significant relation between the model-implied and empirical measures of attention and risky investment share. Furthermore, we provide evidence that attention significantly predicts the VIX. Then, we test the model's predictions that attention is a U-shaped function of the predictive variable, an increasing function of both the absolute predictive coefficient and uncertainty, and a decreasing function of stock-return variance. We show that the data lend support to these theoretical predictions.

In Section 3.1, we used three state variables to calibrate the parameters of the model: the S&P 500 earnings-to-price ratio y_t, the time-varying predictive coefficient $\widehat{\beta}_t$, and the conditional variance of returns V_t. Figure 5 illustrates the dynamics of these state variables, where the gray shaded areas represent NBER recessions.

In order to obtain *model-implied* time series of attention and uncertainty, we first discretize the

	Attention a_t^*	Risky share w_t^*	Uncertainty ν_t
Intercept	0.037***	0.288***	2.085***
	(16.38)	(5.00)	(1208.20)
Recession dummy	0.019***	−0.255***	−0.015***
	(3.30)	(−2.60)	(−2.76)
R^2	0.029	0.009	0.028
Observations	720	720	720

Table 3: Model-implied attention, risky investment share, and uncertainty in NBER recessions.
a_t^*, w_t^*, and ν_t are the model-implied measures of attention, risky investment share, and uncertainty, respectively. Newey and West (1987) t-statistics are reported in brackets and statistical significance at the 10%, 5%, and 1% levels is labeled with */**/***, respectively. Data are at monthly frequency from 01/1955 to 12/2014.

dynamics of uncertainty in (16) as follows:

$$\nu_{t+\Delta} = \nu_t + \left[-\left(\frac{(y_t - \bar{y})^2}{V_t} + a_t^*(\widehat{\mu}_t, y_t, V_t, \nu_t) \right) \nu_t^2 - 2\lambda_\beta \nu_t + \sigma_\beta^2 \right] \Delta, \tag{38}$$

where $\Delta = 1/12 = 1$ month, $\widehat{\mu}_t \equiv \bar{\mu} + \widehat{\beta}_t(y_t - \bar{y})$, the initial value is $\nu_0 = \bar{\nu} = \sigma_\beta^2/(2\lambda_\beta)$, and the optimal attention $a_t^*(\widehat{\mu}_t, y_t, V_t, \nu_t)$ defined in (34) depends on the function $\phi(.)$ defined in (31). The parameter values provided in Table 1 and the solution method described in Appendix B yield the function $\phi(.)$. Therefore, using the initial value $\nu_0 = \bar{\nu}$ and sequentially substituting the values of the state variables depicted in Figure 5 in Equations (33), (34), and (38) provides model-implied time series for the risky investment share, attention, and uncertainty. These model-implied time series are depicted in Figure 6.

We study the behavior of these model-implied quantities over the business cycle using NBER recession dummies. Table 3 shows that the model-implied attention is larger in recessions than in expansions. In recessions, the earnings-to-price ratio spikes and the predictive coefficient drops to negative values to reflect negative expected returns. Investors optimally react to these changes by paying more attention to news (as our theoretical model predicts in Figure 1). Attention is therefore counter-cyclical, consistent with Andrei and Hasler (2014). Table 3 further shows that investors place a smaller fraction of their wealth in the stock in recessions than in expansions, and also that the model-implied uncertainty about the predictive coefficient is pro-cyclical, in line with our theoretical prediction that higher attention tends to decrease uncertainty (see Equation (16)).

It is important to check whether the model-implied measures of attention, uncertainty, and risky

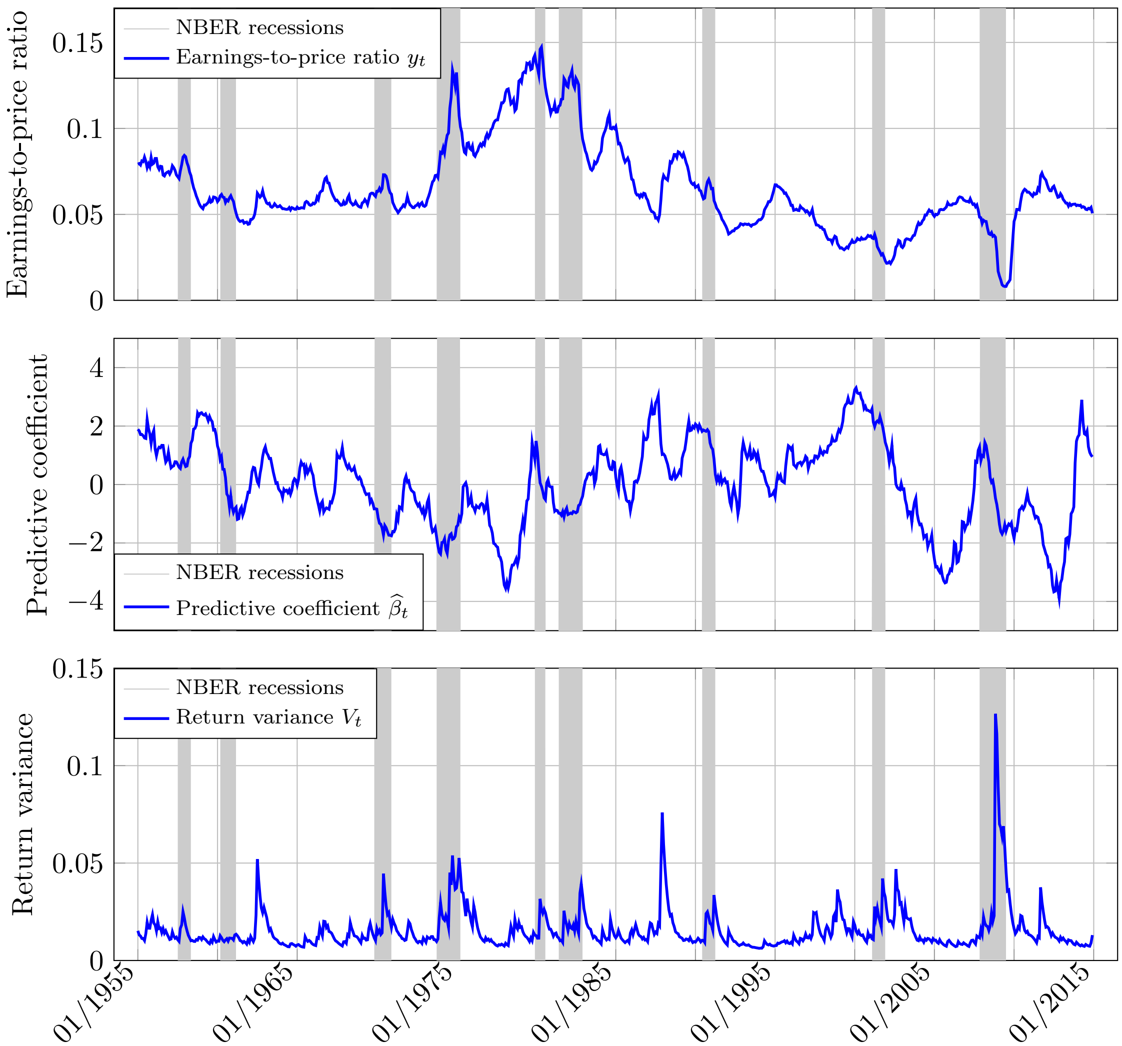


Figure 5: Historical dynamics of the predictive variable, the predictive coefficient, and the stock-return variance.

Data are monthly frequency from 01/1955 to 12/2014.

investment share are realistic, and thus we compare them to their corresponding empirical proxies. To construct the empirical measure of attention, a_t^E, we select firms belonging to the Thomson-Reuters institutional database (13F) that *i)* have a stock price larger than \$5, *ii)* are older than one year, *iii)* have a share code of 10 or 11 (i.e. U.S. firms), *iv)* have a market capitalization above \$20 million, and *v)* have a trading volume larger than 100,000 shares per year. For each selected firm, we obtain the quarterly *1)* trading volume, *2)* number of institutional owners, *3)* number of earnings per share (EPS) forecasts, and *4)* number of EPS forecast revisions. The average number of firms that satisfied the five aforementioned conditions each quarter from Q4/1983 to Q4/2014 is equal to 2,123.

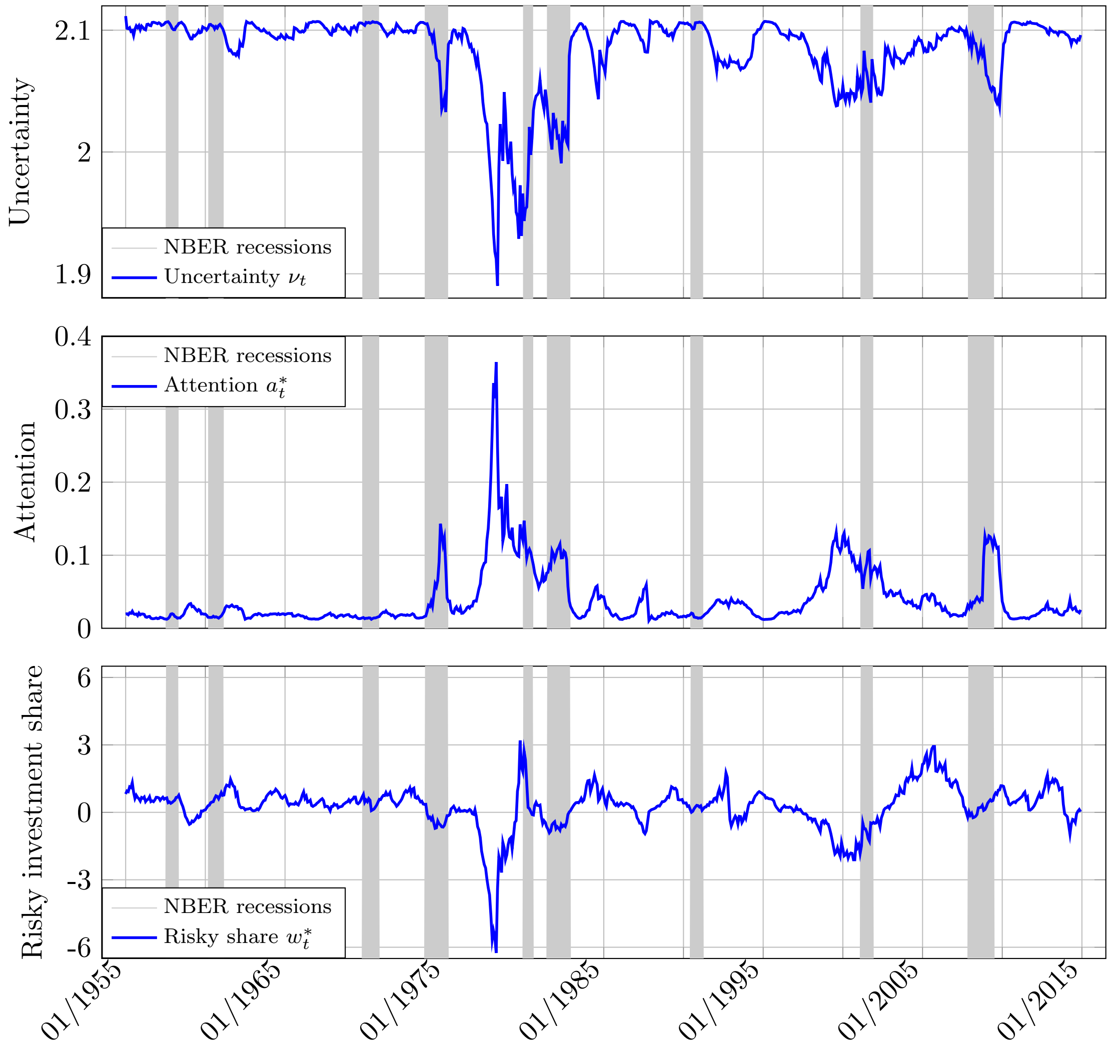


Figure 6: Historical dynamics of model-implied uncertainty, attention, and risky investment share.

Data are monthly frequency from 01/1955 to 12/2014.

Aggregating these firm-specific attention measures using June 30^{th} market capitalizations provides four aggregate non-stationary attention measures. We take the logarithm of these measures, remove their linear time trends, standardize them for scaling purposes, and aggregate them using equal weights. This provides our stationary empirical measure of investors' attention. This time series is at quarterly frequency from Q4/1983 to Q4/2014.

Our choice of the four attention variables defined above is motivated by the empirical literature. First, Chordia and Swaminathan (2000), Lo and Wang (2000), Gervais, Kaniel, and Mingelgrin (2001), Barber and Odean (2008), and Hou, Peng, and Xiong (2009) argue that trading volume is a reasonable proxy for attention because investors must trade stocks they pay attention to. In addition, Fisher et al. (2016) provide evidence that media attention is closely related to trading volume. Second, Boone and White (2015) show that the larger the number of institutional owners, the higher the information production about the firm. That is, more institutional owners attracts more attention. Third, Womack (1996) and Irvine (2004) show that analysts' recommendations and forecasts yield abnormal trading volumes. This suggests that more analyst coverage triggers more attention. Finally, as argued by Jacob, Lys, and Neale (1999), the number of forecast revisions measures analysts' attention to recent news on the firm.

To construct the empirical measure of uncertainty, η_t^E, we use the 1-month-ahead macro-uncertainty index constructed by Jurado et al. (2015). This time series is at monthly frequency from 08/1960 to 12/2014. Note that this macro-uncertainty index is a measure of macroeconomic risk, whereas in our model ν_t is the Bayesian uncertainty regarding the predictive coefficient β_t, which measures learning inaccuracy rather than macroeconomic uncertainty. In our framework, a measure of economic uncertainty comparable with the Jurado et al. (2015) index is the uncertainty about expected returns η_t, which we define in Equation (20) of Lemma 1.

We proxy the empirical risky investment share, w_t^E by *the negative of* the log growth rate of the "Institutional Money Funds" index. The "Institutional Money Funds" index represents the dollar amount held by institutions in the money market. This index is at monthly frequency from 10/1978 to 12/2014, and it is obtained from the Federal Reserve Bank of St. Louis.

We then regress the empirical measures of attention, uncertainty, and risky investment share on their model-implied counterparts. Table 4 shows the results. There is a strong positive relation between the empirical and model-implied measures of attention, even after controlling for the autocorrelation in the empirical measure. This suggests that our model provides a realistic description of the dynamic attention allocation problem faced by investors. Furthermore, Table 4 shows a strong positive relation between the empirical and model-implied measures of uncertainty, although the relation loses significance after controlling for the autocorrelation in the empirical measure. As in Jurado et al. (2015), our model-implied uncertainty about expected returns is counter-cyclical. The correlation coefficient between the model-implied measure and the empirical measure is 0.6.[22] Finally, Table 4 shows that the empirical and model-implied measures of the risky investment share

[22] In unreported results, we also find a strong positive correlation (0.52) between the model-implied measure of uncertainty about expected returns and the VIX index.

	Attention a_t^E	Attention a_t^E	Uncertainty η_t^E	Uncertainty η_t^E	Risky share w_t^E	Risky share w_t^E
Intercept	-0.370***	-0.077**	0.636***	0.015	-0.017***	-0.007***
	(-2.77)	(-2.03)	(152.1)	(1.37)	(-7.55)	(-4.25)
a_t^*	10.77***	2.29**				
	(3.18)	(2.00)				
$a_{t-3/12}^E$		0.819***				
		(16.23)				
η_t^*			25.19***	0.908		
			(10.5)	(1.43)		
$\eta_{t-1/12}^E$				0.976***		
				(55.9)		
w_t^*					0.007***	0.003***
					(3.21)	(2.63)
$w_{t-1/12}^E$						0.598***
						(11.22)
R^2	0.274	0.782	0.359	0.978	0.039	0.385
Observations	125	124	653	652	435	434

Table 4: Empirical vs. model-implied attention, uncertainty and risky investment share. The variables a_t^E, η_t^E, and w_t^E denote the empirical measures of attention, uncertainty and risky investment share, respectively. The variables a_t^*, η_t^*, and w_t^* denote the model-implied counterparts. We report Newey and West (1987) t-statistics in brackets and label statistical significance at the 10%, 5%, and 1% levels with */**/***. Uncertainty and attention data are at monthly frequency from 08/1960 to 12/2014 and at the quarterly frequency from Q4/1983 to Q4/2014. The monthly model-implied attention a_t^* is averaged over three consecutive months to obtain a quarterly measure. Risky investment share data are at monthly frequency from 10/1978 to 12/2014.

are strongly positively related, even after controlling for the autocorrelation in the empirical measure. This result suggests that, in addition to accurately describing investors' dynamic attention behavior, our model also explains the dynamics of institutional investors' risky investments.

We further test whether the model-implied and empirical measures of attention predict the VIX index. As we show in Equation (21), attention is a proxy for the variance of expected returns, which should imply a positive relation between current attention and future VIX. Table 5 shows that this is indeed the case. Both the model-implied and empirical measures of attention positively predict the VIX, even after controlling for the autocorrelation in the latter. That is, attention can be interpreted as a measure of future market risk. This is also consistent with recent findings by Fisher et al. (2016), who document that an increase in media attention positively relates to an increase in implied volatility.

	VIX_t	VIX_t	VIX_t	VIX_t
Intercept	0.153***	0.029***	0.196***	0.059***
	(12.70)	(5.06)	(18.48)	(7.31)
$a^*_{t-1/12}$	1.123***	0.148*		
	(4.59)	(1.95)		
$\text{VIX}_{t-1/12}$		0.823***		0.700***
		(20.03)		(20.77)
$a^E_{t-1/12}$			0.040***	0.008**
			(7.42)	(2.24)
R^2	0.201	0.729	0.107	0.530
Observations	299	299	99	99

Table 5: The predictive power of attention on the VIX.
The variables a^E_t and a^*_t denote the empirical and model-implied measures of attention, respectively. We report Newey and West (1987) t-statistics in brackets and label statistical significance at the 10%, 5%, and 1% levels with */**/***. Model-implied and empirical attention data are at monthly frequency from 01/1990 to 12/2014 and at quarterly frequency from Q1/1990 to Q4/2014, respectively.

4.1 Testing the predictions of the model

Our theoretical model predicts that investors' attention is a U-shaped function of the predictive variable, a decreasing function of stock-return variance, and an increasing function of uncertainty. Furthermore, the risky investment share increases with the Sharpe ratio of the stock and features a negative hedging demand, which depends on uncertainty, the predictive variable, and the stock-return variance. In what follows, we describe and test the model's predictions.

Prediction 1. *Equation (34) and the results depicted in Figure 1 show that attention can be approximated as follows:*

$$a_t \approx C_0 + C_1\nu_t^2 + \tilde{C}_2(\widehat{\beta}, V_t)\nu_t^2(y_t - \bar{y})^2 \approx C_0 + C_1\nu_t^2 + C_2\frac{|\widehat{\beta}_t|}{V_t}\nu_t^2(y_t - \bar{y})^2, \tag{39}$$

where $C_1 < 0$ and $C_2 > 0$. That is, the curvature of the U-shaped relation between attention and the demeaned predictive variable decreases with the stock-return variance and increases with both the absolute predictive coefficient and the squared uncertainty. In addition, the squared uncertainty determines the level of the U-shaped relation.

Prediction 2. *Equation (33) and the results depicted in Figure 4 show that the risky investment* *share can be approximated as follows:*

$$w_t \approx K_0 + K_1 \frac{\bar{\mu} + \widehat{\beta}_t(y_t - \bar{y}) - r^f}{V_t} + K_2 \frac{\nu_t(y_t - \bar{y})^2}{V_t}, \tag{40}$$

where $K_1 > 0$ and $K_2 < 0$. That is, the risky investment share increases with the Sharpe ratio of the stock and features a negative hedging demand that is hump-shaped in the demeaned predictive variable. The curvature of the hedging demand decreases with the stock-return variance and increases with uncertainty.

The coefficients C_0, C_1, C_2, K_0, K_1, and K_2 are estimated by ordinary least squares using model-implied data to quantify the significance of these coefficients in our model on the one hand, and using the empirical proxies to test the model's predictions on the other hand. The left-hand side of Equations (39) and (40) define the dependent variables, while the right-hand side define the independent variables.

The first and second columns of Table 6 quantify Prediction 1 using model-implied attention data. As expected, the coefficients C_1 and C_2 are negative and positive, respectively, and both are highly statistically significant. Comparing the R^2 obtained in these two columns shows that adjusting the level of the quadratic relation between attention and the predictive variable with the squared uncertainty only has a marginal impact on the goodness of fit. The third and fourth columns test Prediction 1 using the empirical measure of attention. Consistent with the prediction of the model, the curvature of the quadratic relation between attention and the predictive variable decreases with stock-return variance and increases with both the absolute value of the predictive coefficient and the squared uncertainty. In addition, the fourth column shows that, as for its model-implied counterpart, the empirical attention decreases with the squared uncertainty. The coefficient, however, is statistically insignificant, which implies that the R^2 obtained in the fourth column is only marginally larger than that obtained in the third column. This is in line with the model's prediction that adding the squared uncertainty as an independent variable does not have a significant impact on the goodness of fit.

While we provide theoretical and empirical evidence of a quadratic relation between attention and the earnings-to-price ratio, Fisher et al. (2016) show that there also exists a quadratic relation between media attention and macroeconomic fundamentals.

	Attention a_t^*	Attention a_t^*	Attention a_t^E	Attention a_t^E
Intercept	0.019***	0.920***	-0.180	9.908
	(16.94)	(6.18)	(-1.62)	(1.40)
ν_t^2		-0.205***		-2.292
		(-6.08)		(-1.41)
$\frac{\lvert\widehat{\beta}_t\rvert}{V_t}\nu_t^2(y_t-\bar{y})^2$	0.114***	0.053***	1.41***	0.692***
	(20.63)	(4.72)	(3.40)	(3.33)
R^2	0.713	0.810	0.247	0.273
Observations	375	375	125	125

Table 6: Model-implied and empirical measures of attention vs. predictive variable, predictive coefficient, return variance, and uncertainty.
The variables a_t^E and a_t^* are the empirical and model-implied measures of attention, respectively. We report Newey and West (1987) t-statistics in brackets and label statistical significance at the 10%, 5%, and 1% levels with */**/***. Model-implied and empirical attention data are at monthly frequency from 10/1983 to 12/2014 and at quarterly frequency from Q4/1983 to Q4/2014, respectively. In the fourth and fifth columns, the monthly model-implied predictive variable y_t, predictive coefficient $\widehat{\beta}_t$, return variance V_t, and uncertainty ν_t are averaged over three consecutive months to obtain quarterly measures.

Using model-implied data, the first and second columns of Table 7 quantify Prediction 2 and confirm that the coefficients K_1 and K_2 are positive and negative, respectively. That is, the risky investment share increases with the Sharpe ratio of the stock and features a negative hedging demand. The second column also shows that the hedging demand is statistically insignificant in our model. The third and fourth columns of Table 7 test Prediction 2 using the empirical proxy for the risky investment share. Consistent with the model's prediction, the empirical measure of the risky investment share increases with the Sharpe ratio of the stock and features a negative hedging demand that is hump-shaped in the predictive variable. Furthermore, the curvature of the hump-shaped relation between the hedging demand and the predictive variable decreases with stock-return variance and increases with uncertainty. Comparing the significance of the model-implied hedging term (the second column of Table 7) with the significance of the empirical hedging term (the fourth column of Table 7) shows that our model tends to underestimate the hedging motives of institutional investors.

Finally, the model also provides a prediction on the conditional relation between attention and the risky investment share. Indeed, Figures 1 and 4 suggest that large positive values of the product $\widehat{\beta}_t(y_t-\bar{y})$ imply both high attention and a large positive risky investment share. Intuitively, when expected returns are high, investors are highly attentive and increase their risky investment share.

	Risky share w_t^*	Risky share w_t^*	Risky share w_t^E	Risky share w_t^E
Intercept	0.001	0.002	−0.017***	−0.007***
	(0.13)	(0.09)	(−7.53)	(−3.91)
$\frac{\bar{\mu}+\widehat{\beta}_t(y_t-\bar{y})-r^f}{V_t}$	0.337***	0.337***	0.003***	0.002***
	(48.51)	(45.68)	(2.83)	(3.09)
$\frac{\nu_t(y_t-\bar{y})^2}{V_t}$		−0.002		−0.083***
		(−0.01)		(−5.39)
R^2	0.970	0.970	0.044	0.217
Observations	435	435	435	435

Table 7: Model-implied and empirical risky investment share vs. Sharpe ratio and hedging demand.
The variables w_t^E and w_t^* are the empirical and model-implied measures of risky investment share, respectively. We report Newey and West (1987) t-statistics in brackets and label statistical significance at the 10%, 5%, and 1% levels with */**/***. Data are at monthly frequency from 10/1978 to 12/2014.

In contrast, large negative values of the product $\widehat{\beta}_t(y_t-\bar{y})$ imply high attention and a large negative risky investment share (according again to Figures 1 and 4). This provides the testable Prediction 3 described below.

Prediction 3. *The model predicts a positive relation between he risky investment share and attention when $\widehat{\beta}_t(y_t-\bar{y})$ is highly positive, and a negative relation when $\widehat{\beta}_t(y_t-\bar{y})$ is highly negative. That is, the relation between the risky investment share and attention satisfies*

$$w_t = M_0 + M_1\widehat{\beta}_t(y_t-\bar{y})a_t, \tag{41}$$

where $M_1 > 0$.

The first column of Table 8 quantifies Prediction 3 and shows that the coefficient K_1 is, as expected, highly significant in the model. Importantly, the second column of Table 8 tests the prediction using our empirical proxies and shows that it finds empirical support.

5 Conclusion

This paper aims at understanding the dynamic attention behavior observed in financial markets. In most of the existing literature, investors acquire information passively in the sense that they do not control the quality of information they collect. In contrast, we consider an investor who can,

	Risky share w_t^*	Risky share w_t^E
Intercept	0.464***	-0.008***
	(3.38)	(-2.95)
$\widehat{\beta}_t(y_t - \bar{y})a_t^*$	198.50***	
	(3.94)	
$\widehat{\beta}_t(y_t - \bar{y})a_t^E$		0.175***
		(4.41)
R^2	0.488	0.075
Observations	375	125

Table 8: Model-implied and empirical relation between risky investment share and attention.
The variables a_t^E and a_t^* are the empirical and model-implied measures of attention, respectively. The variables w_t^E and w_t^* are the empirical and model-implied measures of risky investment share, respectively. We report Newey and West (1987) t-statistics in brackets and label statistical significance at the 10%, 5%, and 1% levels with */**/***. The second and third columns consider data at monthly frequency from 10/1983 to 12/2014 and at quarterly frequency from Q4/1983 to Q4/2014, respectively. In the third column, the monthly model-implied predictive variable y_t and predictive coefficient $\widehat{\beta}_t$ are averaged over three consecutive months to obtain quarterly measures.

at each point in time, improve the accuracy of acquired information at a cost.

Our analysis provides several interesting insights. The optimal level of attention paid to news is a U-shaped function of the stock return predictor, an increasing function of uncertainty about predictability, a decreasing function of stock-return volatility, and an increasing function of the absolute predictive coefficient. In addition, the risky investment share increases with the Sharpe ratio of the stock and features a negative hedging demand. The hedging demand is a hump-shaped function of the return predictor, an increasing function of stock-return volatility, and a decreasing function of uncertainty. We show that the data lends support to these theoretical predictions, and conclude that empirically documented fluctuations in investors' attention result from a rational information gathering behavior.

Our analysis can be extended to a multiple asset setting, which would help understand the impact of costly dynamic information acquisition on diversification. It would also be interesting to investigate the impact of the optimal choice of attention on the equilibrium risk-free rate, equity premium, and equity return volatility in a pure-exchange economy. Finally, a production economy can offer insights on the impact of costly information acquisition on the dynamics of aggregate consumption.

References

Abel, A., J. Eberly, and S. Panageas (2007). Optimal inattention to the stock market. *American Economic Review 97*, 244–249.

Abel, A. B., J. C. Eberly, and S. Panageas (2013). Optimal inattention to the stock market with information costs and transactions costs. *Econometrica 81*(4), 1455–1481.

Ait-Sahalia, Y. and M. W. Brandt (2001). Variable selection for portfolio choice. *The Journal of Finance 56*(4), 1297–1351.

Andrei, D., B. Carlin, and M. Hasler (2017). Asset pricing with structural uncertainty and structural disagreement. Technical report, UCLA.

Andrei, D. and M. Hasler (2014). Investor attention and stock market volatility. *The Review of Financial Studies 28*(1), 33.

Andries, M. and V. Haddad (2014). Information aversion. *Working Paper*.

Barber, B. M. and T. Odean (2008). All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *Review of Financial Studies 21*(2), 785–818.

Barberis, N. (2000, 02). Investing for the long run when returns are predictable. *Journal of Finance 55*(1), 225–264.

Barberis, N., M. Huang, and T. Santos (2001, February). Prospect theory and asset prices. *The Quarterly Journal of Economics 116*(1), 1–53.

Bloom, N. (2009, 05). The impact of uncertainty shocks. *Econometrica 77*(3), 623–685.

Bollerslev, T. (1986, April). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics 31*(3), 307–327.

Boone, A. L. and J. T. White (2015). The effect of institutional ownership on firm transparency and information production. *Journal of Financial Economics 117*(3), 508–533.

Brandt, M. W. (1999, October). Estimating portfolio and consumption choice: A conditional euler equations approach. *Journal of Finance 54*(5), 1609–1645.

Brandt, M. W., A. Goyal, P. Santa-Clara, and J. R. Stroud (2005). A simulation approach to dynamic portfolio choice with an application to learning about return predictability. *Review of Financial Studies 18*(3), 831–873.

Brennan, M. J. (1998). The role of learning in dynamic portfolio decisions. *European Finance Review 1*(3), 295–306.

Brennan, M. J. and Y. Xia (2001). Stock price volatility and equity premium. *Journal of Monetary Economics 47*(2), 249 – 283.

Brown, A. L. and H. Kim (2013). Do individuals have preferences used in macro-finance models? an experimental investigation. *Management Science 60*(4), 939–958.

Brunnermeier, M. K. and J. A. Parker (2005). Optimal expectations. *The American Economic Review 95*, 1092–1118.

Campbell, J. Y. and J. H. Cochrane (1999). By force of habit: A consumption-based explanation of aggregate stock market beha. *Journal of Political Economy 107*, 205–251.

Campbell, J. Y. and L. M. Viceira (1999, May). Consumption and portfolio decisions when expected returns are time varying. *The Quarterly Journal of Economics 114*(2), 433–495.

Caplin, A. and J. Leahy (2001). Psychological expected utility theory and anticipatory feelings. *Quarterly Journal of economics*, 55–79.

Caplin, A. and J. Leahy (2004). The supply of information by a concerned expert*. *The Economic Journal 114*(497), 487–505.

Chacko, G. and L. M. Viceira (2005). Dynamic consumption and portfolio choice with stochastic volatility in incomplete markets. *Review of Financial Studies 18*(4), 1369–1402.

Chien, Y., H. Cole, and H. Lustig (2012, September). Is the volatility of the market price of risk due to intermittent portfolio rebalancing? *American Economic Review 102*(6), 2859–96.

Chordia, T. and B. Swaminathan (2000). Trading volume and cross-autocorrelations in stock returns. *The Journal of Finance 55*(2), 913–935.

Cochrane, J. H. (2008). The dog that did not bark: A defense of return predictability. *Review of Financial Studies 21*(4), 1533–1575.

Cohn, A., J. Engelmann, E. Fehr, and M. A. Maréchal (2015). Evidence for countercyclical risk aversion: an experiment with financial professionals. *The American Economic Review 105*(2), 860–885.

Da, Z., J. Engelberg, and P. Gao (2011, October). In search of attention. *Journal of Finance 66*(5), 1461–1499.

Das, S. R. and R. Uppal (2004). Systemic risk and international portfolio choice. *The Journal of Finance 59*(6), 2809–2834.

David, A. (1997, December). Fluctuating confidence in stock markets: Implications for returns and volatility. *Journal of Financial and Quantitative Analysis 32*(04), 427–462.

Detemple, J. (1986). Asset pricing in a production economy with incomplete information. *The Journal of Finance 41*(2), pp. 383–391.

Detemple, J. and R. Kihlstrom (1987). Acquisition d'information dans un modèle intertemporel en temps continu. *L'Actualité économique 63*, 118–137.

Detemple, J. B., R. Garcia, and M. Rindisbacher (2003). A monte carlo method for optimal portfolios. *The Journal of Finance 58*(1), pp. 401–446.

Dothan, M. U. and D. Feldman (1986, June). Equilibrium interest rates and multiperiod bonds in a partially observable economy. *Journal of Finance 41*(2), 369–82.

Eichenbaum, M. S., L. P. Hansen, and K. J. Singleton (1988). A time series analysis of representative agent models of consumption and leisure choice under uncertainty. *The Quarterly Journal of Economics 103*(1), 51–78.

Epstein, L. G., E. Farhi, and T. Strzalecki (2014). How much would you pay to resolve long-run risk? *The American Economic Review 104*(9), 2680–2697.

Feldman, D. (1989). The term structure of interest rates in a partially observable economy. *The Journal of Finance 44*(3), pp. 789–812.

Fisher, A., C. Martineau, and J. Sheng (2016). Media attention, macroeconomic fundamentals, and stock market activity. *Working Paper*.

Galai, D. and O. Sade (2006). The "Ostrich Effect" and the Relationship between the Liquidity and the Yields of Financial Assets. *The Journal of Business 79*(5), 2741–2759.

Gennotte, G. (1986, July). Optimal portfolio choice under incomplete information. *Journal of Finance 41*(3), 733–46.

Gervais, S., R. Kaniel, and D. H. Mingelgrin (2001). The high-volume return premium. *The Journal of Finance 56*(3), 877–919.

Goyal, A. and I. Welch (2008, July). A Comprehensive Look at The Empirical Performance of Equity Premium Prediction. *Review of Financial Studies 21*(4), 1455–1508.

Hall, R. E. (1988). Intertemporal substitution in consumption. *Journal of political economy 96*(2), 339–357.

Hansen, L. P. and K. J. Singleton (1983). Stochastic consumption, risk aversion, and the temporal behavior of asset returns. *Journal of political economy 91*(2), 249–265.

Hasler, M. and C. Ornthanalai (2017). Fluctuating attention and financial contagion. *Working Paper*.

Heston, S. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies 6*(2), 327–343.

Hou, K., L. Peng, and W. Xiong (2009). A tale of two anomalies: The implications of investor attention for price and earnings momentum. *Working Paper*.

Huang, L. and H. Liu (2007, 08). Rational inattention and portfolio selection. *Journal of Finance 62*(4), 1999–2040.

Irvine, P. J. (2004). Analysts' forecasts and brokerage-rirm trading. *The Accounting Review 79*(1), 125–149.

Jacob, J., T. Z. Lys, and M. A. Neale (1999). Expertise in forecasting performance of security analysts. *Journal of Accounting and Economics 28*(1), 51 – 82.

Judd, K. L. (1998). *Numerical Methods in Economics.* MIT Press.

Jurado, K., S. C. Ludvigson, and S. Ng (2015). Measuring uncertainty. *American Economic Review 105*(3), 1177–1216.

Kacperczyk, M., S. Van Nieuwerburgh, and L. Veldkamp (2016). A rational theory of mutual funds' attention allocation. *Econometrica 84*(2), 571–626.

Kandel, S. and R. F. Stambaugh (1996). On the predictability of stock returns: An asset-allocation perspective. *The Journal of Finance 51*(2), pp. 385–424.

Karlsson, N., G. Loewenstein, and D. Seppi (2009, April). The ostrich effect: Selective attention to information. *Journal of Risk and Uncertainty 38*(2), 95–115.

Kim, T. S. and E. Omberg (1996). Dynamic nonmyopic portfolio behavior. *Review of Financial Studies 9*(1), 141–61.

Liptser, R. S. and A. N. Shiryaev (2001). *Statistics of Random Processes II.* Springer Verlag, New York.

Liu, J. (2007). Portfolio selection in stochastic environments. *Review of Financial Studies 20*(1), 1–39.

Liu, J., E. Peleg, and A. Subrahmanyam (2010, October). Information, expected utility, and portfolio choice. *Journal of Financial and Quantitative Analysis 45*(05), 1221–1251.

Lo, A. and J. Wang (2000). Trading volume: definitions, data analysis, and implications of portfolio theory. *Review of Financial Studies 13*(2), 257–300.

Loewenstein, G. (1987). Anticipation and the valuation of delayed consumption. *The Economic Journal 97*(387), 666–684.

Merton, R. C. (1971, December). Optimum consumption and portfolio rules in a continuous-time model. *Journal of Economic Theory 3*(4), 373–413.

Merton, R. C. (1998). *Continuous-time finance.* Malden, Massachusetts [etc]: Blackwell.

Mondria, J. (2010). Portfolio choice, attention allocation, and price comovement. *Journal of Economic Theory 145*(5), 1837 – 1864.

Mondria, J. and C. Quintana-Domeque (2013). Financial contagion and attention allocation. *The Economic Journal 123*(568), 429–454.

Newey, W. K. and K. D. West (1987, May). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*(3), 703–08.

Pagel, M. (2013). A news-utility theory for inattention and rebalancing in portfolio choice. Technical report, Working paper.

Peng, L. and W. Xiong (2006). Investor attention, overconfidence and category learning. *Journal of Financial Economics 80*(3), 563 – 602.

Sicherman, N., G. Loewenstein, D. J. Seppi, and S. P. Utkus (2015). Financial attention. *The Review of Financial Studies 29*(4), 863.

Sims, C. A. (2003, April). Implications of rational inattention. *Journal of Monetary Economics 50*(3), 665–690.

Stambaugh, R. F. (1999). Predictive regressions. *Journal of Financial Economics 54*(3), 375–421.

Thornton, R. L. (2008). The demand for, and impact of, learning hiv status. *The American economic review 98*(5), 1829–1863.

Van Nieuwerburgh, S. and L. Veldkamp (2006). Learning asymmetries in real business cycles. *Journal of Monetary Economics 53(4)*, 753–772.

Van Nieuwerburgh, S. and L. Veldkamp (2010). Information acquisition and under-diversification. *Review of Economic Studies 77(2)*, 779–805.

Veldkamp, L. (2006a). Media frenzies in markets for financial information. *American Economic Review 96(3)*, 577–601.

Veldkamp, L. L. (2006b). Information markets and the comovement of asset prices. *Review of Economic Studies 73*(3), 823–845.

Veronesi, P. (1999). Stock market overreaction to bad news in good times: A rational expectations equilibrium model. *Review of Financial Studies 12*(5), 975–1007.

Veronesi, P. (2000, April). How does information quality affect stock returns? *Journal of Finance 55*(2), 807–837.

Wachter, J. A. (2002). Portfolio and consumption decisions under mean-reverting returns: An exact solution for complete markets. *The Journal of Financial and Quantitative Analysis 37*(1), pp. 63–91.

Womack, K. L. (1996). Do brokerage analysts' recommendations have investment value? *The Journal of Finance 51*(1), 137–167.

Xia, Y. (2001). Learning about predictability: The effects of parameter uncertainty on dynamic asset allocation. *The Journal of Finance 56*(1), pp. 205–246.

Ziegler, A. C. (2003). *Incomplete Information and Heterogeneous Beliefs in Continuous-time Finance*. Springer-Verlag Berlin Heidelberg.

Appendix

A Proof of Proposition 2

The dynamics of the vector of state variables $Z_t \equiv [\widehat{\mu}_t, y_t, V_t, \nu_t]^\top$ satisfy

$$dZ_t = m_t dt + \Sigma_{1,t} d\widehat{B}^{\perp}_{1,t} + \Sigma_{2,t} \begin{pmatrix} d\widehat{B}^{\perp}_{2,t} \\ d\widehat{B}^{\perp}_{3,t} \\ d\widehat{B}^{\perp}_{4,t} \end{pmatrix}, \tag{42}$$

where the 4-dimensional vector of drift m, the 4-dimensional vector of diffusion Σ_1, and the 4×3 matrix of diffusion Σ_2 satisfy

$$m = \begin{pmatrix} (\lambda_y + \lambda_\beta)\left(\bar{\mu} + \frac{\bar{\beta}\lambda_\beta (y-\bar{y})}{\lambda_y + \lambda_\beta} - \widehat{\mu}\right) \\ \lambda_y(\bar{y} - y) \\ \lambda_V(\bar{V} - V) \\ -\left(\frac{(y-\bar{y})^2}{V} + a\right)\nu^2 - 2\lambda_\beta \nu + \sigma_\beta^2 \end{pmatrix} \qquad \Sigma_1 = \begin{pmatrix} \frac{(y-\bar{y})^2 \nu}{\sqrt{V}} \\ 0 \\ 0 \\ 0 \end{pmatrix} \tag{43}$$

$$\Sigma_2 = \begin{pmatrix} \sigma_y \frac{\widehat{\mu} - \bar{\mu}}{y - \bar{y}} & 0 & \nu\sqrt{a}(y - \bar{y}) \\ \sigma_y & 0 & 0 \\ 0 & \sigma_V \sqrt{V} & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{44}$$

The HJB equation satisfies

$$0 = -\delta J + \max_{c,a,w} \left(u(c_t) + \mathcal{D}^{W,Z} J\right), \tag{45}$$

where $\mathcal{D}^{W,Z}$ is the infinitesimal generator such that[23]

$$\mathcal{D}^{W,Z} J = J_Z' m + J_W \left[(r^f - K(a_t))W_t + w_t W_t \left(\widehat{\mu}_t - r^f\right) - c_t\right] \tag{46}$$

$$+ \frac{1}{2} J_{WW} W_t^2 w_t^2 V_t + \frac{1}{2} \mathrm{tr}\left[(\Sigma_1 \Sigma_1' + \Sigma_2 \Sigma_2') J_{ZZ}\right] \tag{47}$$

$$+ W_t w_t \sqrt{V_t} \Sigma_1' J_{WZ}. \tag{48}$$

Differentiating (23) partially with respect to the control variables yields the first order conditions:

$$0 = u_c - J_W \tag{49}$$

$$0 = J_W W_t (\widehat{\mu}_t - r^f) + J_{WW} W_t^2 V_t w_t + J_{W\mu} W_t \nu_t (y_t - \bar{y})^2 \tag{50}$$

$$0 = -K'(a_t) J_W W_t - \nu_t^2 J_\nu + \frac{1}{2} \nu_t^2 (y_t - \bar{y})^2 J_{\mu\mu}. \tag{51}$$

Solving the first order conditions yields Proposition 2.

[23] Note that, for notational convenience, we drop hats when state variables appear as indices.

B Numerical solution method

B.1 Chebyshev collocation

Substituting $K(x) = kx^2$, $u(x) = \frac{x^{1-\gamma}}{1-\gamma}$, and the first order conditions (49)-(51) in Equation (45) yields a partial differential equation (PDE) for the value function $J(W, \widehat{\mu}, y, V, \nu)$. Then, using the conjecture $J(W, \widehat{\mu}, y, V, \nu) = \frac{W^{1-\gamma}}{1-\gamma}\phi(\widehat{\mu}, y, V, \nu)$ provides a PDE for the function $\phi(\widehat{\mu}, y, V, \nu)$.

Since the state variables $\widehat{\mu}_t$ and y_t belong to the real line by definition, they do not imply boundary conditions. In contrast, the dynamics of V_t and ν_t imply that $V = 0$ and $\nu = 0$ can be approached but not attained, implying two boundary conditions.

When $\nu = 0$, the investor observes the predictive coefficient β_t and, thus, the expected return μ_t, $\forall t$. This immediately implies that $\widehat{\mu}_t \equiv \mu_t$ and $a_t = K(a_t) = 0$, $\forall t$. Writing the HJB equation using the dynamics of W_t, μ_t, y_t, and V_t, substituting the first order conditions, and conjecturing the solution

$$J_{|\nu=0}(W, \mu, y, V) = \frac{W^{1-\gamma}}{1-\gamma}\phi_{|\nu=0}(\mu, y, V), \tag{52}$$

yields a 3-dimensional PDE for $\phi_{|\nu=0}(\mu, y, V)$. By definition, the solution $J_{|\nu=0}(W, \mu, y, V)$ is the boundary condition at $\nu = 0$.

Setting $V = 0$ implies that the stock becomes riskless. By no arbitrage, the expected return μ_t must be equal to the risk-free rate r^f, $\forall t$. That is, the value function $J_{|V=0}(W)$ at the boundary/corner $V = 0$ only depends on wealth and is therefore straightforward to compute.

The PDE for $\phi(\widehat{\mu}, y, V, \nu)$ is solved numerically using the Chebyshev collocation method (Judd, 1998). That is, we approximate the function $\phi(\widehat{\mu}, y, V, \nu)$ as follows:

$$\phi(\widehat{\mu}, y, V, \nu) \approx P\left(\widehat{\mu}, y, V, \nu\right) = \sum_{i=0}^{I}\sum_{j=0}^{J}\sum_{k=0}^{K}\sum_{l=0}^{L} a_{i,j,k,l} T_i\left(\widehat{\mu}\right) T_j\left(y\right) T_k\left(V\right) T_l\left(\nu\right),$$

where T_m is the Chebyshev polynomial of order m. The interpolation nodes are obtained by meshing the scaled roots of the Chebyshev polynomials of order $I+1$, $J+1$, $K+1$, and $L+1$. We scale the roots of the Chebyshev polynomials of order $I+1$, $J+1$, $K+1$, and $L+1$ such that they cover the intervals $\widehat{\mu} \in [q_{\widehat{\mu},1\%}, q_{\widehat{\mu},99\%}]$, $y \in [q_{y,1\%}, q_{y,99\%}]$, $V \in [0, q_{V,99\%}]$, and $\nu \in [0, q_{\nu,99\%}]$, respectively. Note that $q_{x,p\%}$ stands for the p percentile of the process x. The polynomial $P\left(\widehat{\mu}, y, V, \nu\right)$ and its partial derivatives are then substituted into the PDE, and the resulting expression is evaluated at the interpolation nodes. This yields a system of $(I+1) \times (J+1) \times (K+1) \times (L+1)$ equations with $(I+1) \times (J+1) \times (K+1) \times (L+1)$ unknowns (the coefficients $a_{i,j,k,l}$) that is solved numerically. The mean squared PDE residual computed over the set of 180 interpolation nodes is of order 10^{-29}.

B.2 Transversality, feasibility, and optimality conditions

We define the function $F(c, a, w)$ as follows:

$$F(c, a, w) \equiv -\delta J + u(c_t) + \mathcal{D}^{W,Z} J, \tag{53}$$

where $\mathcal{D}^{W,Z}$ is provided in (48). Equation (45) can therefore be rewritten as

$$\max_{c,a,w} F(c, a, w) = 0. \tag{54}$$

As argued in Merton (1998), a solution to problem (23) must satisfy

1. the transversality condition

$$\lim_{t\to\infty} \mathbb{E}\left(e^{-\delta t} J(W_t, \widehat{\mu}_t, y_t, V_t, \nu_t)\right) = 0, \tag{55}$$

2. the sufficient condition for a maximum

$$\text{Hessian of } F(c,a,w) \equiv \begin{pmatrix} F_{cc} & F_{cw} & F_{ca} \\ F_{wc} & F_{ww} & F_{wa} \\ F_{ac} & F_{aw} & F_{aa} \end{pmatrix} \text{ has only negative eigenvalues,} \tag{56}$$

3. and the two feasibility conditions

   - $c_t^* \geq 0$
   - $a_t^* \geq 0$.

The first order conditions of (54) are

$$F_c = 0 = u_c - J_W \tag{57}$$

$$F_w = 0 = J_W W_t(\widehat{\mu}_t - r^f) + J_{WW} W_t^2 V_t w_t + J_{W\mu} W_t \nu_t (y_t - \bar{y})^2 \tag{58}$$

$$F_a = 0 = -K'(a_t) J_W W_t - \nu_t^2 J_\nu + \frac{1}{2}\nu_t^2 (y_t - \bar{y})^2 J_{\mu\mu}. \tag{59}$$

From this, we immediately obtain

$$F_{cw} = F_{wc} = F_{ca} = F_{ac} = F_{wa} = F_{aw} = 0, \qquad\qquad F_{cc} = u_{cc} < 0 \tag{60}$$

because the utility function $u(.)$ is concave by definition. Furthermore, the value function satisfies

$$J(W_t, \widehat{\mu}_t, y_t, V_t, \nu_t) = \frac{W_t^{1-\gamma}}{1-\gamma}\phi(\widehat{\mu}_t, y_t, V_t, \nu_t). \tag{61}$$

Therefore, we have

$$F_{ww} = J_{WW} W_t^2 V_t = -\gamma W_t^{1-\gamma}\phi(\widehat{\mu}_t, y_t, V_t, \nu_t) V_t, \tag{62}$$

$$F_{aa} = -K''(a_t) J_W W_t = -2k W_t^{1-\gamma}\phi(\widehat{\mu}_t, y_t, V_t, \nu_t). \tag{63}$$

Since the Hessian of $F(c,a,w)$ is a diagonal matrix, the diagonal elements are the eigenvalues. This implies that the sufficient condition for a maximum is satisfied if $\phi(\widehat{\mu}_t, y_t, V_t, \nu_t) > 0$ (see Equations (62) and (63)). $\phi(\widehat{\mu}_t, y_t, V_t, \nu_t) > 0$ also implies that the first feasibility condition is satisfied (see Equation (32)).

Simulations show that both $\phi(\widehat{\mu}_t, y_t, V_t, \nu_t)$ and attention a_t^* are strictly positive, with minimums equal to about $8,000$ and 1%, respectively. Therefore, the two feasibility conditions and the sufficient condition for a maximum are satisfied. In Figure 7, we depict the behavior of $\mathbb{E}\left(e^{-\delta t} J(W_t, \widehat{\mu}_t, y_t, V_t, \nu_t)\right)$ with respect to time t, which we obtain via simulations. This figure shows that the transversality condition is also satisfied, and therefore that our solution indeed solves the investor's maximization problem (23).

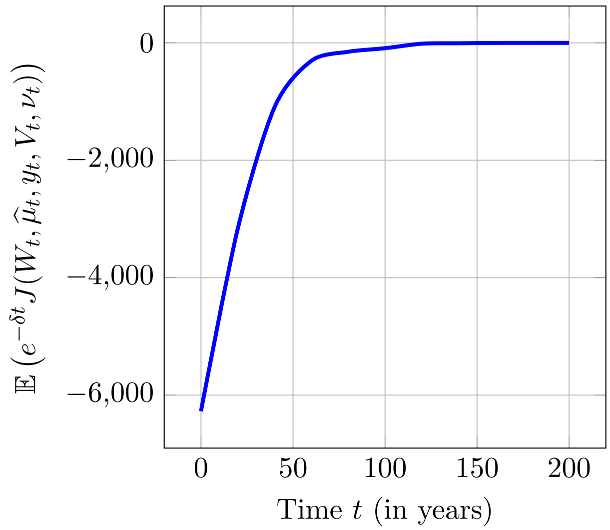


Figure 7: Transversality condition.
Initial state variables are $W_0 = 1$, $\widehat{\beta}_0 = \bar{\beta} = 0$, $y_0 = \bar{y} = 6.9\%$, $\sqrt{V_0} = \sqrt{\bar{V}} = 12.6\%$, and $\nu_0 = \bar{\nu} = 2.11$. Parameter values are provided in Table 1.

C Alternative information cost specification

Let us assume the following more general information cost function

$$\overline{K}_t = k_1 a_t^{k_2} W_t^{k_3}, \tag{64}$$

where $k_1 > 0$ and $k_2 > 1$. Setting $k_1 = k$, $k_2 = 2$, and $k_3 = 1$ yields the quadratic form specification from Section 2.4.

The general cost function (64) is *ex ante* independent of wealth if $k_3 = 0$, decreasing with wealth if $k_3 < 0$ (i.e. information gets cheaper as time passes), and increasing with wealth if $k_3 > 0$ (i.e. information gets more expensive as time passes). Note that k_2 has to be larger that 1 in order to obtain a maximum (see first equality in Equation (63)).

At the optimum, we obtain

$$a_t^* = \left(\frac{1}{2k_1k_2}\right)^{1/(k_2-1)} \left(\frac{\nu_t^2\left((y_t-\bar{y})^2 J_{\mu\mu} - 2J_\nu\right)}{J_W}\right)^{1/(k_2-1)} W_t^{-k_3/(k_2-1)}, \tag{65}$$

$$\overline{K}_t^* = k_1\left(\frac{1}{2k_1k_2}\right)^{k_2/(k_2-1)} \left(\frac{\nu_t^2\left((y_t-\bar{y})^2 J_{\mu\mu} - 2J_\nu\right)}{J_W}\right)^{k_2/(k_2-1)} W_t^{-k_3/(k_2-1)}. \tag{66}$$

We are interested in the effect of the parameter k_3 on our qualitative results. In order to do so, we approximate (65) by taking a first order Taylor expansion around $k_{11} = 0$, where the parameter k_{11} is defined as

$$k_{11} \equiv \left(\frac{1}{2k_1k_2}\right)^{1/(k_2-1)}, \tag{67}$$

and thus does not depend on k_3. The parameter k_{11} converges to zero as $k_1 \to \infty$. In this case, the cost of information is infinite and therefore the investor does not pay attention to news i.e. the news signal s_t has zero precision. The first order Taylor expansion of attention around $k_{11} = 0$ yields

$$a_t^* \approx \left(\frac{\nu_t^2 W_t^{-k_3} \left((y_t - \bar{y})^2 J_{\mu\mu} - 2J_\nu \right)}{J_W} \right)^{1/(k_2-1)} \Bigg|_{k_{11}=0} \times k_{11}. \tag{68}$$

When the investor does not pay attention to news ($k_{11} = 0$), we know that the value function is homogeneous in wealth and thus it takes the following functional form

$$J(W_t, \widehat{\mu}_t, y_t, V_t, \nu_t) = \frac{W_t^{1-\gamma}}{1-\gamma} G(\widehat{\mu}_t, y_t, V_t, \nu_t), \tag{69}$$

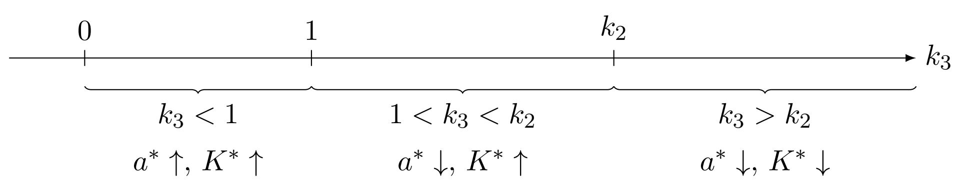
where the function $G(.)$ solves a 4-dimensional PDE. Substituting this in the first term of (68), the optimal attention and the *ex post* cost of information are

$$a_t^* = k_{11} H(\widehat{\mu}_t, y_t, V_t, \nu_t)^{1/(k_2-1)} W_t^{(1-k_3)/(k_2-1)}, \tag{70}$$

$$\overline{K}_t^* = k_1 k_{11}^{k_2} H(\widehat{\mu}, y, V, \nu)^{k_2/(k_2-1)} W_t^{(k_2-k_3)/(k_2-1)} \tag{71}$$

for some function $H(.)$.

This implies that both attention and the *ex post* cost of information increase with wealth if $k_2 > k_3$ and $k_3 < 1$, and decrease with wealth if $k_2 < k_3$. If $k_2 > k_3$ and $k_3 > 1$, then the cost of information increases with wealth, whereas attention decreases with it. The figure below illustrates these three cases.



For a fixed amount of wealth, attention described here and attention described in the main body of the paper are driven by the state variables in the same way. The reason is that the signs and sensitivities of the partial derivatives J_W, $J_{\mu\mu}$, and J_ν are qualitatively the same here and in the main body of the paper (refer to the end of Section 2.3 for a discussion on these partial derivatives). We expect therefore that the qualitative implications of our paper remain unchanged (given a fixed amount fo wealth) under a more general cost function specification such as (64).